High-Resolution Imaging of Earth's Lowermost Mantle

by

Hongyu Lai

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved July 2019 by the
Graduate Supervisory Committee:

Edward Garnero, Chair
Sang-Heon Shim
Mingming Li
Christy Till
James Tyburczy

ARIZONA STATE UNIVERSITY

August 2019

ABSTRACT

This research investigates the fine scale structure in Earth`s mantle, especially for the lowermost mantle, where strong heterogeneity exists. Recent seismic tomography models have resolved large-scale features in the lower mantle, such as the large low shear velocity provinces (LLSVPs). However, differences are present between different models, especially at shorter length scales. Fine scale structures both within and outside LLSVPs are still poorly constrained. The drastic growth of global seismic networks presents densely sampled seismic data in unprecedented quality and quantity. In this work, the Empirical Wavelet construction method has been developed to document seismic travel time and waveform information for a global shear wave seismic dataset. A dataset of 250K high-quality seismic records with comprehensive measurements is documented and made publicly available. To more accurately classify high quality seismic signal from the noise, 1.4 million manually labeled seismic records have been used to train a supervised classification model. The constructed model performed better than the empirical model deployed in the Empirical Wavelet method, with 87% in precision and 83% in recall. To utilize lower amplitude phases such as higher multiples of S and ScS waves, we have developed a geographic bin stacking method to improve signal-to-noise ratio. It is then applied to S_n waves up to $n = 6$ and ScS_n wave up to $n = 5$ for both minor and major arc phases. The virtual stations constructed provide unique path sampling and coverage, vastly improving sampling in the Southern Hemisphere. With the high-quality dataset we have gathered, ray-based layer stripping iterative forward tomography is implemented to update a starting tomography model by mapping the travel time residuals along the ray from the surface down to the core mantle

boundary. Final updated models with different starting tomography models show consistent updates, suggesting a convergent solution. The final updated models show higher resolution results than the starting tomography models, especially on intermediate-scale structures. The combined analyses and results in this work provide new tools and new datasets to image the fine-scale heterogeneous structures in the lower mantle, which advances our understanding of the dynamics and evolution of the Earth`s mantle.

DEDICATION

I dedicate this dissertation to my supporting and loving parents, who have supported me all along the way. I would not be able to make this far without them, their open-mind and peaceful heart.

I also dedicate this work to my advisor Ed Garnero, who have been my mentor and my friend throughout my 6 years of life in Arizona. His everlasting passion and creativity inspired me to explore my own path and chase for the ‘impact’. The perspective and intuition I have learnt from him helped me to become a thorough scientist as well as a more thoughtful person. With the never settling spirit I inherited from him, I would continue to explore, create and grow. And that is my greatest gratitude towards him.

Finally, I dedicate this dissertation to my coffee infused Arizona morning, my late-night reflection and meditation, my unforgettable journeys across the United States, my exciting exploration and growth in artificial intelligence, and my splendid School of Earth and Space Exploration.

ACKNOWLEDGEMENT

I want to express my gratitude towards my committee members, Dan Shim, Mingming Li, Christy Till and James Tyburczy, for their considerate and thorough comments and advice on doing better science.

I want to thank my academic brother, Shule Yu, who has shared the same path with me and has helped me at untenable moments and crisis times. I would also like to thank my colleagues, Divya Allu Peddinti, John West, Carole Nisr, Huawei Chen, Byeongkwan Ko and Harrison Allen-Sutter for the friendly and supportive lab life. And also thanks to my friends in SESE, Alyssa Anderson, Alana Williams, Heather Meyer, Kyle Rogers, Jessica Noviello and Meghan Guild, who have made my graduate school life so much more enjoyable.

I want to thank my dear friends, Tianyi Shen, Shuyao Hong, Abigail Weibel, Xiaoda Sun, Devon Lichtenstein, Seth Levine, Mike Anastasi, Kaiqing Yuan, Sunyoung Park, Feifei Zhang, who have shared so much of my unforgettable life in Arizona and I am grateful for the support and companion of Mengyao Cheng and Hélène Piet.

My special thanks goes to my therapist, Brian Blocker, who has helped me to learn a great deal about myself.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

The format and content of each chapter conform to the author guidelines of American Geophysical Union.

CHAPTER 1

1 INTRODUCTION

1.1 Introduction

Earth`s lower mantle is rich in heterogeneity. At longest wave length, two antipodal large low shear velocity provinces (LLSVPs) exist right above the core mantle boundary (CMB) [*Masters and Laske*, 2000; *Mégnin and Romanowicz*, 2000; *Grand*, 2002; *Ritsema et al.*, 2011], one beneath the Pacific Ocean and the other beneath the Africa and the southern Atlantic Ocean. The large-scale structures of LLSVPs have been consistently identified from different global seismic tomography models from different groups [*Lekic et al.*, 2012; *Garnero et al.*, 2016].

However, the intermediate to small scale features differ between different models. Smaller scale (as small as several hundred km lateral extent at depth of 2800 km) heterogeneities are observed both within and outside of LLSVPs in the lower mantle. Outside of LLSVPs, small patches of low velocity provinces exist beneath Perm, Russia [*Lekic et al.*, 2012] identified through clustering analysis and travel time analysis. Local full waveform tomography study shows intermediate sized low velocity zones beneath Northern Pacific Ocean [*Suzuki et al.*, 2016]. Forward modeling studies have shown that isolated structures at scales of 1000 km exist at the base of the Pacific Ocean [*He and Wen*, 2012]. At smaller scale, thin (as thin as 10 km) patches with strong wave speed reductions right above the CMB are observed, which are referred to as Ultralow Velocity Zones (ULVZs). Their locations seem to correlate with the margins of LLSVPs [*McNamara et al.*, 2010a]. Additional small-scale heterogeneity at the scale of 1-10 km have been identified [*Frost et al.*, 2013], suggesting strong scattering effects near LLSVP

margins. At the edge of LLSVPs, sharp gradients of shear velocity have been shown in seismic tomography [*Garnero et al.*, 2016]. Seismic waveform studies that analyze the effect of waveform broadening caused by seismic wave multipathing also suggest sharp interfaces [*Ni*, 2002; *Sun et al.*, 2007; 2009], with locations that coincide with those identified in seismic tomography. Using plate tectonics history, the reconstructed location of large igneous provinces strongly correlates with the present day LLSVP edges[*Burke et al.*, 2008; *Torsvik et al.*, 2010]. The vertical extension of LLSVPs edges as well as the tops of LLSVPs have also been characterized [*Sun et al.*, 2007; *He and Wen*, 2012; *Lekic et al.*, 2012; *Zhao et al.*, 2015]. And within LLSVPs, a variety of heterogeneities have been identified from seismic studies. A local shear velocity discontinuity has been identified that is associated with the post-perovskite phase change [*Lay et al.*, 2006], strong lateral shear velocity variations have been identified at various locations [*Garnero and Helmberger*, 1996; *Bréger and Romanowicz*, 1998; *Russell et al.*, 1998; *He and Wen*, 2009], shear wave anisotropy is observed both within and outside of LLSVPs [*Ritsema et al.*, 1998; *Russell et al.*, 1999; *Moore et al.*, 2004] and ULVZs are identified from different studies from both inside and outside of LLSVPs as well [*Mori and Helmberger*, 1995; *Kohler et al.*, 1997; *Revenaugh and Meyer*, 1997; *Thorne and Garnero*, 2004; *Avants et al.*, 2006; *Hutko et al.*, 2009; *McNamara et al.*, 2010b].The heterogeneities in the lower mantle, including the intermediate sized low velocity structure outside of LLSVPs, the scattered ULVZs that are both within and outside of LLSVPs, the internal variations of LLSVPs and the small-scale scattering near LLSVP edges, all suggest a rich and complex dynamics system in the lower mantle that is highly coupled with LLSVPs.

And detailed structure and morphology of LLSVPs provide constraints on the rheology and dynamics of the lower mantle.

Our understanding of lower mantle heterogeneity has been greatly improved during the past 2 decades, especially with the advancement of global seismic tomography. Simultaneous inversion of seismic data and geodynamic observations using mineral parameters have been carried out to resolve both V_P and V_S [*Simmons et al.*, 2010]. A large collection of Rayleigh wave phase velocity, teleseismic body-wave travel-time and normal mode splitting function measurements have been used to construction a degree 40 shear velocity model [*Ritsema et al.*, 2011]. Independent determination of V_P and V_S have been implemented on long wavelength structures emphasizing the anti-correlation between shear and bulk velocity speed in the lower mantle [*Koelemeijer et al.*, 2016]. And recent development of full waveform tomography incorporating accurate wavefield simulation provides a powerful way to accurately utilize seismic waveform information [*French and Romanowicz*, 2015; *Durand et al.*, 2017]. Analysis of different seismic tomography models provides confidence about the large-scale structures [*Lekic et al.*, 2012; *Garnero et al.*, 2016]. However, the intermediate to small scale features differ in both amplitude and scale, due to the different quality and type of data used, the theoretical assumptions and simplifications, different model parameterization and different degrees of normalization and smoothing.

Publicly available global seismic networks continue to grow, with thousands of seismic stations (temporary or permanent) available for any given earthquake. Data agencies, such as the Incorporated Research Institutions for Seismology (IRIS), F-net Broadband Seismograph Network (F-NET), Northern California Earthquake Data Center

(NCEDC), Canadian National Seismograph Network (CNSN) and Observatories & Research Facilities for European Seismology (ORFEUS) provide freely available digital waveform seismic data in large volume and global seismic datasets continue to rapidly grow and play an important role in Earth structure determination in both forward and inverse studies. The deployment of high quality modern seismic networks brings the opportunity to study Earth`s lower mantle in unprecedented ways with the availability of large volume high quality seismic data among which is the USArray that is carried out in the EarthScope project. More than 700 high quality broadband seismic stations are deployed across the United States of which 400 are deployed in a rolling grid that advances every 2 years. These seismic networks present some of the highest quality seismic waveform data and the volume continues to grow during the program`s 15 years lifetime. At the same time, the large volume of seismic data presents challenges on how to retrieve reliable information among the vast amount of raw seismic data in a way that is systematic and objective.

In this work, a comprehensive data set of global earthquakes is collected. To avoid the effect of shallow structure on the earthquake side, events with source depth greater than 50km from 1994 to 2017 are collected. A total of 822 events are collected with a time window from the source origin time to 2 hours after. As deep events tend to occur at subduction zones, most deep events provide very similar event-station geometry. To diversify the geometry, 578 shallow events are also collected. In the end, a total amount of more than 1400 events are documented for transverse, radial and vertical components of motion, with a total of 7 Terabytes of time series seismic waveform data. The work presented in this dissertation focuses on the transverse component of the shear

wave (V_{SH}). In Chapter 2, we have designed a semi-automated seismic information documentation system that is able to systematically and objectively document seismic measurements, including travel time and waveform information. This method uses iteratively stacked S waves of each event as a reference waveform (Empirical Wavelet). During the stacking, the attenuation effect is corrected by normalizing the waveform width. And the final Empirical Wavelet is then adapted to fit the waveform of individual records for each given phase. The travel time is further defined on a Gaussian function that best fits the adapted Empirical Wavelet. During this process, comprehensive measurements are documented for each individual record, including signal-to-noise ratio, cross-correlation coefficients, waveform misfit between record and adapted Empirical wavelet, Gaussian factor of the best fit Gaussian function, t* factor that describes the attenuation degree et al. The method is applied to 360 events for phases including S, Sdiff, SS, SSS, ScS and ScSScS. A total number of 1.4 million seismic records have been processed this way and after manual confirmation, a dataset of ~250K high quality seismic data entries have been documented and made publicly available.

To assist our manual confirmation process in Empirical Wavelet construction, an empirical classification algorithm is implemented to distinguish high quality and noisy seismic records based on rules associated with signal-to-noise ratio, cross-correlation coefficients and waveform misfit measurements. However, during practice the performance of the empirical classification algorithm is not satisfied, and misclassification happens when the signal-to-noise ratio drops to around 2.0 or the waveform of the data gets complicated. Distinguishing seismic signal from the background noise has been a challenging task and is usually implemented through

manual efforts (e.g. documentation of seismic data archive in Bulletin of the International Seismological Centre, documentation of USArray seismic data by Array Network Facility [*Astiz et al.*, 2014]). A better classification model that is able to accurately distinguish high quality data from the noise is highly desired, which combined with the automated Empirical Wavelet construction would allow fully automated seismic data characterization and information retrieval. The supervised machine learning model learns a function by mapping the input to the output labels based on the example input-output pairs [*Russell & Norvig, 2010*]. It has been successfully implemented in many applications that involve large and complex datasets, including computer vision, natural language processing and data analytics. After exploring the correlation of the comprehensive measurements that we have documented in Empirical Wavelet construction and the target labels (high quality versus low quality), a binary classification model is trained on the Empirical Wavelet dataset in Chapter 3. A fast, distributed, high performance gradient boosting framework (LightGBM) is used during the training and boosting and bagging have been implemented in the training process to reduce bias and avoid overfitting. The final model is tested in the reserved dataset and scored a precision of 87% and recall of 83%, which is a significant improvement compared to the empirical classification model.

From the Empirical Wavelet construction, 80% of the processed seismic records have been identified as noisy and rejected from the final dataset, mostly due to their low signal-to-noise ratio. In Chapter 4, a stacking method is designed to construct virtual stations that stacks seismic records from neighboring stations. Geographical bin stacking has been used in a forward study [*Avants et al.*, 2006] and a seismic array study [*Rost*,

2002]. Being able to stack neighboring stations allows for significant signal-to-noise ratio boosting, which makes seismic data that are noisy distinguishable and also provides the potential to apply the Empirical Wavelet construction method to multiple bounce phases that are usually noisy, in this study referred as multi-bounce phases (Sn for n up to 6 and ScSn for n up to 5). As the phases bounce multiple times in the mantle, both minor path (with great arc distance smaller than 180 degree) and major path (with great arc distance greater than 180 degree) phases become available. The virtual stacking method is applied to the 6-basic phases (S, Sdiff, SS, SSS, ScS and ScSScS) as well as the multi-bounce phases. And combined with the methodology designed in the Empirical Wavelet construction, about 9K high quality seismic data entries have been identified from a total of 250K virtual stacks. The coverage provided by this new dataset shows a significant boost in coverage, especially in Southern hemisphere.

And finally in Chapter 5, we utilize the seismic measurements documented from previous experiments and developed a novel layer stripping iterative forward tomography method to construct a new 3D seismic velocity model. This method updates a starting model iteratively from the surface to the core mantle boundary. And after the top layer update, the top layer is kept frozen from further update and the subsequent update is carried out on the next layer. This iterative forward approach is designed to restore the amplitude and pattern of seismic shear velocity that is usually heavily normalized and smoothed in seismic tomography inversion. Combined with the high-quality seismic dataset that we have documented, including the quality measurements used as weights, updated seismic models have been constructed with different starting tomography models. We compare the difference in both the starting model and updated models and

we show more consistent features in both pattern and scale in the updated models. The consistency in updated models also shows independency from starting models and provide confidence that iterative forward tomography is updating starting models into a converged solution. We have further identified several locations with low velocity anomaly in the updated models, which are consistent with forward and inverse studies on the same regions. Additionally, strong velocity perturbation is observed within LLSVPs, similar to the pattern predicted from geodynamic models [*Garnero and McNamara*, 2008], suggesting the thermochemical nature of LLSVPs.

This work aims to improve our understanding of lower mantle structures through incorporation of large volume and high-quality seismic data. The automated approach as mentioned in Chapter 2 allows systematical and objective documentation of seismic information, including travel-time and waveform information. The information that we have documented are shared publicly and can be used in forward modeling and inversion for various kind of studies. The results presented in Chapter 2 are published in Geochemistry, Geophysics and Geosystems (doi:10.1029/2018GC007905). The algorithmic characterization of seismic data, combined with supervised machine learning models as discussed in Chapter 3, opens a new way for studying the patterns and anomaly that we see in seismic data and is applicable to the vast amount of data in modern seismic network. We plan to have the results presented in Chapter 3 to be published in a machine learning conference. In Chapter 4, combined with virtual stacking method to boost the signal-to-noise ratio, the Empirical Wavelet construction method is further expanded to multiple bounce phases, which boosted the data coverage, especially in Southern hemisphere. The results presented in Chapter 4 are ready for submission. And using the

seismic information, we have developed an iterative forward tomography method to construct updated seismic velocity model from starting seismic tomography. In Chapter 5, we show consistent features in both pattern and amplitude in the updated models with different starting models, emphasizing the rich and heterogenous features in the lower mantle. We are in the process of preparing the results presented in Chapter 5 for publication.

1.2 References

Astiz, L. et al. (2014), The Array Network Facility Seismic Bulletin: Products and an Unbiased View of United States Seismicity, *Seismological Research Letters*, *85*(3), 576–593, doi:10.1785/0220130141.

Avants, M., T. Lay, and E. J. Garnero (2006), A new probe of ULVZ S-wave velocity structure: Array stacking of ScS waveforms, *Geophysical Research Letters*, *33*(7), doi:10.1029/2005GL024989.

Bréger, L., and B. Romanowicz (1998), Three-Dimensional Structure at the Base of the Mantle Beneath the Central Pacific, *Science*, *282*(5389), 718–720, doi:10.1126/science.282.5389.718.

Burke, K., B. Steinberger, T. H. Torsvik, and M. A. Smethurst (2008), Plume Generation Zones at the margins of Large Low Shear Velocity Provinces on the core–mantle boundary, *Earth and Planetary Science Letters*, *265*(1-2), 49–60, doi:10.1016/j.epsl.2007.09.042.

Durand, S., E. Debayle, Y. Ricard, C. Zaroli, and S. Lambotte (2017), Confirmation of a change in the global shear velocity pattern at around 1000 km depth, *Geophysical Journal International*, *211*(3), 1628–1639, doi:10.1093/gji/ggx405.

French, S. W., and B. Romanowicz (2015), Broad plumes rooted at the base of the Earth's mantle beneath major hotspots, *Nature*, *525*(7567), 95–99, doi:10.1038/nature14876.

Frost, D. A., S. Rost, N. D. Selby, and G. W. Stuart (2013), Detection of a tall ridge at the core–mantle boundary from scattered PKP energy, *Geophysical Journal International*, *195*(1), 558–574, doi:10.1093/gji/ggt242.

Garnero, E. J., and A. K. McNamara (2008), Structure and Dynamics of Earth's Lower Mantle, *Science*, *320*(5876), 626–628, doi:10.1126/science.1148028.

Garnero, E. J., A. K. McNamara, and S.-H. Shim (2016), Continent-sized anomalous zones with low seismic velocity at the base of Earth's mantle, *Nature Geosci*, *9*(7), 481–489, doi:10.1038/ngeo2733.

Garnero, E. J., and D. V. Helmberger (1996), Seismic detection of a thin laterally varying boundary layer at the base of the mantle beneath the central-Pacific, *Geophysical Research Letters*, *23*(9), 977–980, doi:10.1029/95GL03603.

Grand, S. P. (2002), Mantle shear–wave tomography and the fate of subducted slabs, *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *360*(1800), 2475–2491, doi:10.1098/rsta.2002.1077.

He, Y., and L. Wen (2009), Structural features and shear-velocity structure of the "Pacific Anomaly," *Journal of Geophysical Research: Solid Earth (1978–2012)*, *114*(B2), doi:10.1029/2008JB005814.

He, Y., and L. Wen (2012), Geographic boundary of the "Pacific Anomaly" and its geometry and transitional structure in the north, *Journal of Geophysical Research: Solid Earth (1978–2012)*, *117*(B9), doi:10.1029/2012JB009436.

Hutko, A. R., T. Lay, and J. Revenaugh (2009), Localized double-array stacking analysis of PcP: D″ and ULVZ structure beneath the Cocos plate, Mexico, central Pacific, and north Pacific, *Physics of the Earth and Planetary Interiors*, *173*(1-2), 60–74, doi:10.1016/j.pepi.2008.11.003.

Kohler, M. D., J. E. Vidale, and P. M. Davis (1997), Complex scattering within D″ observed on the very dense Los Angeles Region Seismic Experiment Passive Array, *Geophysical Research Letters*, *24*(15), 1855–1858, doi:10.1029/97GL01823.

Lay, T., J. Hernlund, E. J. Garnero, and M. S. Thorne (2006), A Post-Perovskite Lens and D‘“’ Heat Flux Beneath the Central Pacific, *Science*, *314*(5803), 1272–1276, doi:10.1126/science.1133280.

Lekic, V., S. Cottaar, A. Dziewonski, and B. Romanowicz (2012), Cluster analysis of global lower mantle tomography: A new class of structure and implications for chemical heterogeneity, *Earth and Planetary Science Letters*, *357-358*, 68–77, doi:10.1016/j.epsl.2012.09.014.

Masters, G., and G. Laske (2000), The Relative Behavior of Shear Velocity, Bulk Sound Speed, and Compressional Velocity in the Mantle: Implications for Chemical and Thermal Structure,, 1–28.

McNamara, A. K., E. J. Garnero, and S. Rost (2010a), Tracking deep mantle reservoirs with ultra-low velocity zones, *Earth and Planetary Science Letters*, *299*(1-2), 1–9, doi:10.1016/j.epsl.2010.07.042.

Mégnin, C., and B. Romanowicz (2000), The three-dimensional shear velocity structure of the mantle from the inversion of body, surface and higher-mode waveforms, *Geophysical Journal International*, *143*(3), 709–728, doi:10.1046/j.1365-246X.2000.00298.x.

Moore, M. M., E. J. Garnero, T. Lay, and Q. Williams (2004), Shear wave splitting and waveform complexity for lowermost mantle structures with low-velocity lamellae and transverse isotropy, *J. Geophys. Res.*, *109*(B2), 173–26, doi:10.1029/2003JB002546.

Mori, J., and D. V. Helmberger (1995), Localized boundary layer below the mid-Pacific velocity anomaly identified from a PcP precursor, *Journal of Geophysical Research: Solid Earth (1978–2012)*, *100*(B10), 20359–20365, doi:10.1029/95JB02243.

Ni, S. (2002), Sharp Sides to the African Superplume, *Science*, *296*(5574), 1850–1852, doi:10.1126/science.1070698.

Revenaugh, J., and R. Meyer (1997), Seismic Evidence of Partial Melt Within a Possibly Ubiquitous Low-Velocity Layer at the Base of the Mantle, *Science*, *277*(5326), 670–673, doi:10.1126/science.277.5326.670.

Ritsema, J., A. Deuss, H. J. Van Heijst, and J. H. Woodhouse (2011), S40RTS: a degree-40 shear-velocity model for the mantle from new Rayleigh wave dispersion, teleseismic traveltime and normal-mode splitting function measurements, *Geophysical Journal International*, *184*(3), 1223–1236, doi:10.1111/j.1365-246X.2010.04884.x.

Ritsema, J., T. Lay, E. J. Garnero, and H. Benz (1998), Seismic anisotropy in the lowermost mantle beneath the Pacific, *Geophysical Research Letters*.

Rost, S. (2002), Array seismology: Methods and applications, *Rev. Geophys.*, *40*(3), 1008, doi:10.1029/2000RG000100.

Russell, S. A., T. Lay, and E. J. Garnero (1998), Seismic evidence for small-scale dynamics in the lowermost mantle at the root of the Hawaiian hotspot,, *396*(6708), 255–258, doi:10.1038/24364.

Russell, S. A., T. Lay, and E. J. Garnero (1999), Small-scale lateral shear velocity and anisotropy heterogeneity near the core-mantle boundary beneath the central Pacific imaged using broadband ScS waves, *Journal of Geophysical Research: Solid Earth (1978–2012)*, *104*(B6), 13183–13199, doi:10.1029/1999JB900114.

Russell, S. J., and Norvig, P. (2016). Artificial intelligence: a modern approach. Malaysia; *Pearson Education Limited.*

Simmons, N. A., A. M. Forte, L. Boschi, and S. P. Grand (2010), GyPSuM: A joint tomographic model of mantle density and seismic wave speeds, *J. Geophys. Res.*, *115*(B12), B12310, doi:10.1029/2010JB007631.

Sun, D., D. Helmberger, S. Ni, and D. Bower (2009), Direct measures of lateral velocity variation in the deep Earth, *J. Geophys. Res.*, *114*(B5), B05303, doi:10.1029/2008JB005873.

Sun, D., E. Tan, D. Helmberger, and M. Gurnis (2007), Seismological support for the metastable superplume model, sharp features, and phase changes within the lower mantle, *PNAS*, *104*(22), 9151–9155, doi:10.1073/pnas.0608160104.

Suzuki, Y., K. Kawai, R. J. Geller, A. F. E. Borgeaud, and K. Konishi (2016), Waveform inversion for 3-D S-velocity structure of D″ beneath the Northern Pacific: possible evidence for a remnant slab and a passive plume, *Earth, Planets and Space 2016 68:1*, *68*(1), 198, doi:10.1186/s40623-016-0576-0.

Thorne, M. S., and E. J. Garnero (2004), Inferences on ultralow-velocity zone structure from a global analysis of SPdKSwaves, *J. Geophys. Res.*, *109*(B8), 421–22, doi:10.1029/2004JB003010.

Torsvik, T. H., K. Burke, B. Steinberger, S. J. Webb, and L. D. Ashwal (2010), Diamonds sampled by plumes from the core-mantle boundary, *Nature*, *466*(7304), 352–355, doi:10.1038/nature09216.

Zhao, C., E. J. Garnero, A. K. McNamara, N. Schmerr, and R. W. Carlson (2015), Seismic evidence for a chemically distinct thermochemical reservoir in Earth's deep mantle beneath Hawaii, *Earth and Planetary Science Letters*, *426*(C), 143–153, doi:10.1016/j.epsl.2015.06.012.

CHAPTER 2

2 GLOBAL TRAVEL TIME DATASET WITH EMPIRICAL WAVELET CONSTRUCTION

Abstract

We present a method for constructing the average waveform shape (hereafter called “empirical wavelet”) of seismic shear waves on an event-by-event basis for the purpose of constructing a high-quality travel time dataset with information about waveform quality and shape. A global dataset was assembled from 360 earthquakes between 1994-2017. The empirical wavelet approach permits documentation of the degree of similarity of every observed wave with the empirical wavelet. We adapt the empirical wavelet to all pulse widths, thus identifying broadened (e.g., attenuated) pulses. Several measures of goodness of fit of the empirical wavelet to each record are documented, as well as signal-to-noise ratios, permitting users of the dataset to employ flexible weighting schemes. We demonstrate the approach on transversely polarized SH waves, and build a global travel time dataset for the waves S, SS, SSS, Sdiff, ScS, and ScSScS. Onset arrival times of the waves were determined through a correlation scheme with best-fitting empirical wavelets. Over 250,000 travel times were picked, from over 1.4 million records, all of which were human-checked for accuracy via a PDF catalog file making system. Many events were specifically selected to bolster southern hemisphere coverage. Coverage maps show that, while the northern hemisphere is more densely sampled, the southern hemisphere coverage is robust. The travel time dataset, empirical wavelets, and all measurement metrics are publicly available, and well suited for global tomography, as well as forward modeling experiments.

2.1 Introduction

Freely available global seismic data continue to increase in volume, with data from thousands of seismographic stations available for any given modern earthquake. Data agencies, such as the Incorporated Research Institutions for Seismology (IRIS, http://www.iris.edu), F-net Broadband Seismograph Network (F-NET, http://www.fnet.bosai.go.jp), Northern California Earthquake Data Center (NCEDC, http://www.ncedc.org), Canadian National Seismograph Network (CNSN, http://www.earthquakescanada.nrcan.gc.ca) and Observatories & Research Facilities for European Seismology (ORFEUS, https://www.orfeus-eu.org), are examples of sources of freely available digital waveform data. The growing regional and global seismic datasets afford an increase in sampling coverage of Earth's mantle, and therefore play an important role in Earth structure determination in both forward and inverse studies. For example, seismic body wave travel time information is an essential component of seismic tomography models of global mantle structure [e.g., *Grand*, 2002; *Houser et al.*, 2008; *Ritsema et al.*, 2011; *Durand et al.*, 2017]. We use all freely available data from these networks to construct an up-to-date global travel time dataset.

A number of methods can be used to measure seismic travel times for global mantle structure studies. These include picking by hand, such as with many of the travel times in the voluminous International Seismology Centre (ISC) dataset, which has been commonly used in P wave tomography studies [*Woodhouse and Dziewonski*, 1984; *Inoue et al.*, 1990; *van der Hilst et al.*, 1997]. Hand-picking is subject to human error and inconsistencies, and moderate noise levels can erroneously modulate wave onset times. As the sheer volume of data continues to rapidly grow, this approach becomes less and

less feasible due to the time-consuming nature of hand picking data. A less subjective approach has been to cross-correlate observed seismic waveforms with synthetic seismograms to obtain travel time residuals [e.g., *Ritsema and van Heijst*, 2002; *Rost*, 2002; *Houser et al.*, 2008]. This method has the benefit that it can be automated, with signal-to-noise (SNR) considerations incorporated. However, data from any given earthquake can have variable pulse widths from source directivity and path effects (e.g., attenuation, multi-pathing, scattering), so that some level of low-pass filtering is often employed to equalize observed pulse widths. Using longer periods ensures a more consistent match to synthetic reference pulses. Waveform information (and travel times) sensitive to structure at shorter scales is desired, but may be subject to variable pulse widths in observations. The body-wave travel time datasets used in seismic tomography models have a wide range, e.g., from an upper period limit of ~15 sec (e.g. 15 sec in GyPSuM [*Simmons et al.*, 2010], 15 sec in HMSL [*Houser et al.*, 2008] , 16 sec in S40RTS [*Ritsema et al.*, 2011], 20 sec for S362ANI [*Kustowski et al.*, 2008]).

A third method for obtaining travel times involves correlation based methods between observed phases, like multi-channel cross-correlation [*Vandecar and Crosson*, 1990; *Schaff and Waldhauser*, 2005; *Pavlis and Vernon*, 2010; *Lou and van der Lee*, 2014]. Users can hand pick an onset time in stacks of data made by such correlation schemes [e.g., *Lou and van der Lee*, 2014] to get absolute time anomalies. These types of algorithms have the benefit of inherently including an averaged source-time function effect from unusual source processes (i.e., effects that synthetic seismogram construction may not have included). However, as with the other methods, an onset time determined from a stack of data inherently averages any timing differences from variability in wave-

shape (especially width) due to source directivity, lateral variations in attenuation, and possible multi-pathing in the presence of strong heterogeneity.

Another approach, which accommodates variable pulse widths and shapes, is cluster analysis which groups distinctly different waveforms. For example, Houser et al. [2008] employed an automated cross correlation algorithm on all data with high signal to noise ratio. This process results in a cluster tree from which a cluster level is chosen (to maximize the waveform population and remove poor data) on an event-by-event basis. A user then hand picks an onset time for the average waveform shape of each cluster, using synthetic seismograms as a guide. Thus this method is a bit of a hybrid of hand picking and multichannel correlative schemes. However, if waveform width is smoothly varying (e.g., from azimuthal dependence of directivity), clustering may blur onset time differences if too few clusters are chosen.

In this study, we present a method that builds an average waveform shape of observed data for the purpose of documenting travel times, but with the additional objective of accommodating variable pulse width shape in the data used to make the average shape. This permits the use of a relatively shorter period for the corner of the low pass filter, thus retaining effects from smaller scale structural phenomena. We have developed the method with an aim for building a global dataset of transversely polarized S waves. The following sections present the empirical wavelet development assumptions and methodology, the global dataset, and some basic information about the resulting measured travel times.

2.2 Global Dataset

2.2.1 Data Collection

Our goal in this project is to build a global dataset of SH wave travel times. To achieve this, we collected all available data from several agencies that freely share data (see Table 2.1). We initially collected global broadband seismic data for earthquakes in the time period from January 1994 to October 2017, with moment magnitude greater than 6.0 and reported source depth larger than 50 km. The beginning date of 1994 was chosen based on ample digital data available for each event to confidently build empirical wavelets, described in the next section. This resulted in 733 deep earthquakes being collected. The source magnitude and depth restrictions were implemented to obtain earthquakes with enough energy to be observed globally, and with the depth phases (i.e., upgoing energy from the source that produces surface reflections like sS) arriving later than downward traveling direct waves. All events were inspected for possible contamination from other events which occurred nearby in time. If any energy was apparent, (including from local seismicity), the event was omitted from our catalog (only a small number of events were rejected). Spurious energy was noted as arrivals appearing in data in a systematic fashion (e.g., with distance in record sections, or energy localized in time and distance at a very different frequency content). Events with complex source-time functions, either due to very long duration (e.g., from exceptionally large earthquakes) or complex waveforms containing multiple peaks, as well as events with poor SNR (usually less then 2.0) were rejected.

We initially processed every event from 1994 to 2007. To explore path coverage diversity from different station geometries, we processed most events in 2013-2014. It

became clear that many events were in regions in which we already possessed abundant data (e.g., Fiji-Tonga, South America, and northwest Pacific subduction zones). To further optimize global data coverage, a second stage of event collection involved shallow earthquakes with source depth less than 50 km (again with magnitude greater than 6.0). This netted a much larger number of possible events that motivated a prioritization scheme for the shallow events that (a) favored recent time periods (i.e., after 2006, when more network data were available), and (b) ranked events according to the greatest distance from events already processed in our catalog. This insured the most even coverage of earthquakes for any given amount of data processing time devoted to this project. We ended up with 113 shallow events from 1994 to 2017 in our dataset culled from roughly 1400 events; the discarded events were either poor quality (low SNR) or in duplicate locations of other events. These shallow events play an important role in expanding the global coverage provided by the deep earthquakes. We more exhaustively processed some recent year events (e.g. 2017, all source depths) to ensure possible path coverage improvements from moving network arrays, such as the Transportable Array from EarthScope's USArray (http://www.usarray.org, https://earthquake.usgs.gov). Maps showing the geographical distribution of earthquakes and seismic stations are presented in Figure 2.1. Zoom plots illustrate especially dense station distributions in North America, Europe, East-Asia, and the west coast of South America. Our final dataset amounted to 360 earthquakes (113 shallow, 247 deep), with recordings from 8409 unique seismographic station locations.

2.2.2 Basic Processing

We collected a 2-hour time window length following the earthquake origin time for all available seismic stations (i.e., distances between zero and 180°), for all events. For each station, three components (north-south, east-west, up-down) of broadband data are collected, then the horizontal components were rotated to the great-circle path to obtain the radial (R) and transverse (T) components of motion. Here, we focus on the transverse component as our target is SH polarized S waves. The instrument response for each station was removed through deconvolution using the pole-zero file supplied by the data agency. All data were then band pass filtered in the period range between 16 and 100 seconds. The upper filter corner was chosen after trial-and-error experimentation aimed at finding a balance between retaining the shortest periods (which renders some data unusable because of complex higher frequencies) and retaining the largest number of records (which is achieved by stronger low pass filtering, but at a cost of losing unique waveform information at shorter periods).

2.2.3 Example Event

We show an earthquake that occurred on May 28, 2012 as a representative event to illustrate our main data processing procedures. We first present a full record section profile for roughly an hour of data, over the full distance range (Figure 2.2). Travel time curves of the key seismic SH phases that are present are also displayed. This event is deep focus (depth of 591.1 km), so many depth phases are readily apparent. Our data collection for this event yielded 2164 stations, which is too dense to clearly plot; thus, we summed the records every 1.5° in distance (and plot the number of summed records in the small histogram to the right). This event is typical of well recorded earthquakes in

that the seismic waves S, SS, SSS, Sdiff, ScS, and ScSScS (see globes to the right in Figure 2.2) are commonly visible. Thus, in this paper, we focus on these 6 phases, which provide a relatively good sampling of mantle structure with depth (e.g., see the "All" cross-section in Figure 2.2). ScSScSScS ("ScS3") is also visible for this event, but is less commonly observed across our event collection. In Section 2.3, we introduce our approach for building a representative empirical wavelet for each event, which then permits us to build our travel time dataset for these phases for all our events.

2.3 Empirical Wavelet Construction

In this section we introduce our method for building an average shape of the direct S wave for each earthquake, which we define generally as "empirical wavelet". Our approach accommodates the variable pulse width that is present in the data (for every event), and ultimately assigns a travel time for all phases of interest based upon their onset time.

2.3.1 Empirical Wavelet for the Direct S Wave

We construct an empirical wavelet (EW) for each event by averaging the shape of all S waves at distances greater than 30°, up to the core diffraction distance. The minimum distance is introduced to avoid waveform distortions associated with triplications from the 410 and 660 km discontinuities. We set the maximum distance to be before the onset of core diffraction (as defined by ray theory) to avoid using waves which might be anomalously broadened by the effect of diffraction and/or attenuation (and/or scattering) in the empirical wavelet construction process. As we introduce the steps of the EW making process, we present a small subset of the records from the earthquake shown in Figure 2.2 to demonstrate the data processing procedures in a

sequence of figures. The steps which follow outline an iterative EW making process which ultimately normalizes pulse widths of records used in the final EW, in order to preserve the onset shape of the records. This refined EW shape is then broadened or narrowed to fit every record in the process of identifying onset times in the data. This process is similar to iterative stacking/alignment methods of past studies (e.g., Pavlis and Vernon, 2010), except our method adjusts pulse widths to optimize the average pulse shape, which we describe below.

Step1: Construction of the Zeroth Empirical Wavelet (EW0).

In our first step, before stacking the data, we calculate the predicted polarity of the S wave for every station using the focal mechanism information from the Global CMT database [*Dziewonski et al.*, 1981; *Ekström et al.*, 2012]. Any record with a predicted negative polarity is flipped. All records are then normalized with the maximum amplitude set to unity, then stacked within a 40 sec time window (40 sec for S and ScS; 60 sec for all other phases) centered on the PREM [*Dziewonski and Anderson*, 1981] predicted travel time. This initial, EW0, stack is an initial estimation of the S wave average shape. Figure 2.3a presents the EW0 construction process for the small subset of records from the event of Figure 2.2.

Step 2: Iterative EW Updating to Construct the General Empirical Wavelet (GEW).

Once the EW0 estimate is made, the records are shifted to align with EW0 using cross-correlation (Figure 2.3b). They are then re-stacked with this new alignment to construct an updated EW (EW1 in this first iteration), where each record is weighted before stacking according to its cross-correlation coefficient (CCC) and its signal-to-noise ratio (SNR). The SNR is defined as the ratio of the average S-wave amplitude in

the one period time window of the record centered on the PREM predicted time relative to the noise window defined as the average amplitude in an 80 sec window that ends 20 sec before the predicted S wave time. The weighting scheme was constructed to down-weight the influence of anomalous records to the stack, i.e., records with high noise level or significant dissimilarity to the stack. In our experience, records with SNR below 2.2 and CCC less than 0.6 are too noisy to confidently identify the phase of interest, and are thus excluded in the construction of the EW. The EW construction process is iteratively implemented (e.g., Figures 2.3a-c) until the cross-correlation coefficient between the previous and current EW is greater than 0.95. The iterative updating of EW is shown in Figure 2.3d and illustrates the fairly rapid EW evolution to a stable shape. We refer to the final EW (in Figure 2.3d) as the General Empirical Wavelet (GEW).

Step 3: Varying Pulse Width to Construct the Stretched Empirical Wavelet (GEW).

For all earthquakes, seismic wave pulse widths (e.g., for the direct S wave) are variable due to source directivity, variable attenuation along different paths, and possible structural effects like multi-pathing and scattering. However, in the iterative stacking process of *Step 2* these effects are ignored, and thus records possessing different pulse widths are stacked together. The resulting GEW thus possesses a shape of the onset of the waveform that is an average of variable width waveforms. This smoothing effect will reduce the accuracy and confidence in determination of the onset time if using correlative algorithms that match the GEW to the data. Here we seek to construct an EW with a sharper onset shape that is more representative of each record's onset. To achieve this, we apply a wave-shape matching algorithm that matches each record to the GEW for the sake of building a modified EW. Specifically, every record is perturbed in width (by

stretching or compressing in time) to best correlate with the GEW. A family of modified records (for each observation) is made ranging from a stretching factor of 0.5 (50% narrower) to 10 (ten times wider) with an interval of 0.01 (thus 950 stretching modifications). The stretched record that produces the maximum correlation with the GEW is retained. This waveform stretching perturbation process to best correlate with the GEW is illustrated in Figure 2.4, using a record from the population in Figure 2.3. The best matching stretched record for every observation is then used to make an updated EW. This updated EW, made from stretched records that best correlate with the GEW, better retains wave-shape information, since it minimizes any temporal blurring of the pulse onset. We call this new EW (which is stacked using the same weighting scheme as used in the GEW) the Stretched Empirical Wavelet (SEW). A comparison between the GEW and the SEW for the records of Figure 2.2 is presented in Figure 2.5. Immediately apparent is the far reduced standard deviation around the SEW on the main up and downswings of the S wave. The onset of the wave is sharper, and more easily identified. Next, we describe an approach that utilizes the SEW to best match the raw (unaltered, unstretched) seismic waves to determine arrival times.

2.3.2 Onset Time Determination

The short period end of the period range of our data is 16 sec. At this period, variability in waveform pulse width is present. Here we introduce our method of adapting the SEW shape to best fit every record, then to objectively and automatically identify the onset time of the record. We proceed with velocity recordings to take advantage of sharper pulse onsets relative to displacement records, which helps to reduce uncertainties. The steps that follow are a continuation of *Steps 1-3* outlined in the last subsection.

Step 4: SEW Adaptation to Individual Observations

To systemically and objectively identify the onset of each S wave, we employ an adaptive waveform fitting technique that finds a perturbation of the SEW width that best fits each record. Since the SEW is roughly the average width of the population of S waves for each event, half of the S waves will on average be narrower than the SEW, and the other half of the S waves will have broader pulse widths. We build a collection of perturbations of the SEW that span the width range from 50% more narrow, up to 10 times broader. The narrowing of the SEW is achieved simply by changing the time spacing between data points, then re-interpolating (i.e., a time compression). To account for a broadened observation relative to the SEW, we convolve the SEW with a series of t^* operators [*Futterman*, 1962] to simulate the effect of attenuation. We note that we do not incorporate the time shift associated with t^* operators; we simply utilize the wave shape due to the operator. Up to 2000 width gradations of the SEW are generated for each earthquake, and cross-correlated with observed S waves to determine the optimal compressed or expanded SEW that fits each record. Figure 2.6 shows examples of the adaptive waveform fitting approach for narrow and broad records (Figure 2.6a and 2.6b, respectively). The SEW better matches the observed waveform after being adapted to optimally fit. Adapting the SEW to fit observations allows more confident arrival time estimations over the broad range of observed waveform widths.

Step 5: Best-fit Gaussian Functions for Onset Time Determination

Our goal is to determine the onset time of seismic wave arrivals in an automated fashion. After trial and error, we determined that assigning an onset time to a Gaussian function that best reproduces a record's best-fit SEW produces a more stable result

(across all data) than assigning an onset time to the SEW (the Gaussian function and factor are defined in Section 2.4.2). Our approach is as follows: (a) determine the Gaussian function width that gives the highest CCC with the SEW (from *Step 4*) for every individual record. (b) Define the onset time in the Gaussian function from an empirically determined amplitude level: when the Gaussian peak is set at unity, the onset time is fixed to be where the Gaussian amplitude is 0.01 (i.e., the 1% level) for data that has not significantly broadened. Figure 2.7 presents an example of this process. If the data are broadened beyond the SEW with a Gaussian factor greater than 30, we increase the amplitude in the Gaussian function which inherits the onset time, resulting in a later onset time assignment. This method minimizes uncertainties associated with differences in amplitudes with SEW onsets from different earthquakes, since the amplitude growth of the Gaussian (from zero) is systematic. Thus, the onset time for every record is automatically assigned in this process. We also experimented with automatically assigning onset times to the SEW. However, the precursory energy leading up to the SEW onset was variable, thus there was not an amplitude level associated the SEW onset that was uniform across the event population. We briefly note that if we carried this process through with the GEW instead of the SEW, all arrival time estimates would be earlier (how much would depend on each event), because the GEW has a smoother, more distributed, onset time (Figure 2.5). Doing the same procedure with a synthetic seismogram (stretching, onset assignment, etc.) instead of the SEW is possible, but does not accommodate possible source time function variability we see for some events.

2.3.3 Waveform Misfit Measurement

Records with low SNR have less clear waveform onsets. Also, some records may have complex waveforms from multi-pathing or scattering effects that add wave-shape complexity, such as additional shoulders or double peaks, or even precursory energy. The previous steps yield a best-fitting SEW to every S-wave, for every event that is processed. We thus have a means to document how well each observation compares to the adapted SEW shape (and thus a user of the dataset can choose to retain or omit data with significant precursory or misfit energy). Here we define a misfit measurement, which documents the average difference between the record and the best-fitting SEW as follows:

$$Misfit = \frac{\sum_{i=1}^{n} |A_i^{obs} - A_i^{best\ fit\ SEW}|}{n} \quad (2.1)$$

where A_i^{obs} and $A_i^{best\ fit\ SEW}$ are the amplitudes of the i^{th} points of the observed record and best fitting SEW, respectively, where the records have been pre-normalized to unity at the peak amplitude of the S wave. The numerator is thus the summation of the absolute value of the difference between observed and best fit SEW data points, up to the final data point, *n.* This misfit measurement is made for 5 unique time windows: over one period of the phase of interest, over the same period length in the window before and after the phase of interest, and over one additional period length before and following the windows surrounding the main phase window. Here, a period is defined (on a record to record basis) as the length in time over which the amplitude of the phase of interest exceeds 10% of the peak amplitude of the phase (see Figure 2.8), and thus *n* varies from record to record, depending upon pulse width. The central time window is referred to as the main signal window (gray shaded region in Figure 2.8); the preceding and following

windows ("PRE" and "POST", respectively, in Figure 2.8) document the misfit over time segments where pre- and post-cursors, as well as waveform distortions which may depart from the best-fitting SEW, might be detected. These misfits can also be used in weighting schemes in forward or inverse modeling using the resultant travel times. The S waves presented in Figure 2.8 are both clear and visible, with a relatively small misfit for the S wave signal. However, the noise level is clearly higher for the top record (station WZ04), which is apparent in the PRE misfit values (compared to the lower record). Both records have somewhat moderate POST misfit values that are caused by the negative downswing after the S wave window that is not represented in the best fitting SEW (see also the large standard deviation in this POST time region in Figure 2.5). Nonetheless, the stretched SEW fits both records remarkably well, and a confident onset time is achieved. Misfit measurements for the time window following the main phase of interest can be used to hunt for waveforms broadened by multi-pathing, i.e., additional arrivals from reflections or refractions of strong heterogeneity [e.g., Ni and Helmberger, 2003; To et al., 2005].

We have also computed three unique estimations of the SNR of the S wave. Our first method is based on the average amplitude over the one period window of the S wave divided by the average amplitude of a noise window that precedes the S-wave:

$$SNR_{average\_amp} = \frac{\left\{\sum_{i=1}^{n_S} |A_i^S|\right\}/n_S}{\left\{\sum_{i=1}^{n_N} |A_i^N|\right\}/n_N} \tag{2.2}$$

where A_i^S and A_i^N are the i^{th} point amplitude of the Signal ("*S*", the phase of interest) and Noise ("*N*") windows, respectively, and n_S and n_N are the number of points in Signal and Noise windows, respectively. The signal window length is automatically defined as one period, in the same way as with the main phase misfit window, above. The noise window

was set at 80 sec, initiating 100 sec before the PREM predicted onset time. For later arriving phases (e.g., SS, ScS, etc.), for some distances and source depths, there are other phases in this noise window (e.g., direct S, or depth phases like sS). These are masked out by taking the PREM predicted time for all known "traffic" phases and adding a ±15 sec time window around those times, and only using parts of the noise window around those masked time segments.

The second measure of SNR that we employ documents the maximum peak-to-trough amplitude within one period (where the period, *T*, is determined as discussed above) for the signal, as well as throughout the 80 sec noise window (i.e., the one period peak-to-trough maximum within the noise window is retained). These are then used in the SNR ratio:

$$SNR_{peak-trough} = \frac{\max_{0\rightarrow T}[A^{S}_{max\_peak} - A^{S}_{min\_trough}]}{\max_{0\rightarrow T}[A^{N}_{max\_peak} - A^{N}_{min\_trough}]} \quad (2.3)$$

where the $A^{S}_{max\_peak}$ and $A^{S}_{min\_trough}$ are maximum peak and minimum trough amplitudes, respectively, in the signal (*S*) window (i.e., over 1 period, "0→*T*"). The denominator is the same, except that the superscript is *N*, to signify the search for the maximum peak-to-trough within one period over the 80 sec noise window.

The third measure of SNR simply compares the maximum positive peak amplitude of the wave of interest (here, an S wave, which is predefined to be unity) to the maximum positive peak amplitude in the noise window. This differs from the previous SNR measure in that it is not a peak-to-trough measurement, thus it can document large amplitude offsets over time lengths larger than one period (e.g., long period energy, baseline offsets of the seismogram, etc.). It is simply expressed as:

$$SNR_{max\_peak} = \frac{\max_{T}[A^{S}_{max\_peak}]}{\max_{N\ (80\ sec)}[A^{N}_{max\_peak}]} \tag{2.4}$$

where the $A^{S}_{max\_peak}$ and $A^{N}_{max\_peak}$ are the maximum peak amplitudes in the signal window (thus, over one period) and the noise window (over the entire 80 sec), respectively. The noise time window and all three SNR measures are also presented in Figure 2.8.

2.3.4 Data Quality and Catalog PDFs

Here we describe a scheme that automatically classifies each observation as good or poor, and then a PDF file format catalog of the waveforms with the overlain SEW is made for all observations for that event which retains this classification. Ultimately, the PDF is reviewed by humans to either confirm or reject data quality assignments that the algorithm has made. Several factors are used to determine if the records will be automatically characterized as good or poor. They are (a) $SNR_{average\_amp}$: this must be greater than or equal to 2.1 for S, ScS, and 2.2 for Sdiff, SS, SSS and ScSScS to be characterized as good; (b) CCC between the record and its best-fit SEW: this must be greater than or equal to 0.92 to be classified as good for S,ScS and 0. 94 for Sdiff, SS, SSS and ScSScS; (c) onset time anomaly relative to PREM: we require the onset time anomaly be between −15 sec and +20 sec to be considered good (though the human reviewer can update the assignment and include larger time anomalies); and (d) interfering seismic waves ("traffic"): all other seismic waves must be predicted to arrive outside a ±15 sec window relative to the PREM prediction for the phase of interest for the record to be considered good. If characteristics (a), (b), (c) and (d) are all met, then in the PDF, the data are noted as good by using a box with a red "X" in the box, while data that

do not meet these values have the box unchecked (in a PDF business form fashion). Figure 2.9 shows an example of part of a PDF page that shows the original records, globes to the left displaying path geometry, the best-fit SEW and Gaussian, the onset pick, and many data characteristics printed to the right, along with the boxes which are checked if the code characterized the records as "good". The viewer of the PDF can then uncheck or check records based on whether or not records are well classified. We found this an efficient means to review every single seismogram in our dataset. The above values were chosen empirically after many trial and error efforts at effectively picking onsets on broadband velocity SH data in our bandpass filter (16-100 sec). For the 360 events in our dataset, this human viewing/rechecking process took roughly 10 months.

A few additional procedures were included in the making of PDF catalogs. The predicted polarity of our phases of interest (S, SS, SSS, Sdiff, ScS, and ScSScS) were computed using the Global CMT catalog (http://www.globalcmt.org). If the radiation pattern prediction for any of our observations was in the –0.15 to 0.15 range (where maximum radiation pattern values are ±1.0), we colored the record purple in our plots. The record was also flipped in polarity and colored green. The algorithm fit a best-fitting SEW to each polarity of the record so the reviewer of the data has flexibility to choose the flipped polarity version in case the radiation pattern prediction is incorrect. The PDF catalog checked the "good" box for the best version of the two records, if conditions (a) through (d), above, were met. This procedure was motivated by the observation that phases near the nodes of the radiation pattern are sometimes flipped from the CMT solution for the 10 sec period data. This approach was flexible, and the reviewer could easily reject or modify the code's choice, if necessary. Furthermore, records with

predicted low amplitude were only retained if waveform behavior was similar to records with predicted higher amplitudes. The PDF displays all records with the phase of interest having the same polarity (up), and a red dot indicates if a polarity has been flipped for plotting purposes. We plot all predicted arrival times of traffic energy, so the reviewer can modify the code's choices if necessary, since some phases of interest may have travel time anomalies greater than our ±15 sec window used in (d), above, and interfere with traffic energy. As apparent in Figure 2.9, viewing all records from a given earthquake and a given phase at the same time is powerful – the reviewer can identify if a phase is robust (whether it be the phase of interest, or traffic), by viewing near neighbor stations plotted in the PDF near the record being viewed.

In some cases, the best-fit SEW is well matched to the waveform for the phase of interest, but the onset time that is automatically determined does not capture the onset of the observed pulse. This can happen if the wave has experienced multi-pathing and the front part of the wave has pulled out in front of the SEW; this will result in the assigned Gaussian-derived onset time being in error. The PDF catalog includes a numerical entry box to the left of every trace, where the reviewer can zoom into the observed wave and SEW overlay, determine a time shift, then enter the time shift value into the number entry box. For all observed occurrences of this type of wave behavior, we applied corrections so the reported onset times correspond to the actual wave onsets. Once PDF files were saved and closed, we algorithmically extracted information on all selected records from the PDF file.

The empirical wavelet algorithm and classification approach described above has worked well with most earthquakes and the six seismic phases studied here. However, a

small percentage of records experience phase misidentification or false classification (i.e., the SEW gets aligned with a noise pulse). For this reason, the PDF catalog approach with human reviewing was necessary to ensure the highest quality standards. In our quality control process, the reasons we rejected records were that the phase of interest (a) did not have a clear wave onset; (b) had unexpected large precursory energy immediately preceding the wave onset; (c) had nearby unidentifiable large pulses (i.e. $\pm$100 seconds around the phase of interest), which puts the source of the wave interpreted as the phase of interest in question; (d) has relatively low amplitude compared to the background noise level; and (e) has interfering (traffic) phases within 15 sec of its predicted arrival time. Some events were rejected if their source-time function shape was too complex to be fit with a Gaussian (Step 5, Section 2.3.2), which would yield erroneous onset times in our procedure.

Of the ~1.4 Million records processed, the human checking of the algorithm picks resulted in ~5% of originally rejected picks being added back to the retained data, and ~27% of the originally retained data to be rejected. Many good data were rejected by the algorithm based on our SNR criteria (commonly due to some long period energy far ahead of the phase of interest). Some data that the algorithm selected were subsequently rejected because the onset time was unclear, or in some cases the algorithm selected a noise peak. Thus the human reviewing part of this work was important in catching erroneous results from an automated procedure.

2.3.5 Empirical Wavelets of Other S Phases

The first arriving shear wave is the direct S. It possesses the largest SNR and wave-shape stability, and thus the direct S is the wave most representative of the source

time function of each event. We thus use the SEW from direct S as our reference shape in determination of arrival time, wave shape broadening, and misfit of the other phases (SS, SSS, ScS, ScSScS, and Sdiff). The travel time determination of these other phases is the same as for S, namely, we follow *Steps 4* and *5* of Section 2.3.2 to determine onset times; we follow Section 2.3.3 to estimate misfits and SNR values; and we follow Section 2.3.4 to construct PDF catalogs for human reviewing of the automatic picking choices made by the algorithm. We note that SS is a minmax phase with a $\frac{\pi}{2}$ phase shift relative to direct S [*Choy and Richards*, 1975; *Butler*, 1979]. Here we Hilbert transform SS back to the same phase as S (thus, a $\frac{3\pi}{2}$ phase shift). Also, SSS is phase shifted $\frac{\pi}{2}$ beyond the phase of SS, thus a π phase shift beyond S. We similarly put the SSS wave into the phase of S. This allows us to employ the robust SEW created from the direct S for analyses of SS and SSS waves. Figure 2.10 presents examples from three earthquakes, where the SEW constructed from the S wave was adapted to the phase shifted SS and SSS, as well as Sdiff, ScS, and ScSScS. For each event, ten records are shown for each phase. We note that our algorithm did not select data for shallow events if the depth phase was expected to interfere with the phase of interest. The human inspection phase of measurements allowed flexibility for omitting or keeping data picks, based on the depth phase behavior.

Adapting the direct S wave SEW to fit all the other phases has several practical advantages over developing a separate SEW from each phase of interest. First, the direct S has a better SNR compared to all later arriving phases and is easier to detect since it is a first arriving wave on the transverse component of motion. Also, as we describe in the next section, for any earthquake, usable S waves are far more abundant than the later arriving phases, thus more records are used in the SEW stack. Thus, the S wave generated

SEW stacks are more robust. However, as a post-processing step to measuring all phases with a direct S wave generated SEW, it is a simple step to construct SEWs using the other phases (SS, SSS, Sdiff, ScS, ScSScS) since the good data have already been identified for these later arriving waves. In Figure 2.11 we present the GEW for each of the six phases (left column of each panel) along with the standard deviation. A general (and expected) trend is that the higher S multiple phases have broadened GEWs due to the effects of the attenuating mantle. However, this trend is not obvious with all stacks, especially those with very few records in the stack. The column on the right for each event presents the SEW for each phase that was constructed by using the S wave generated GEW for aligning and shape-adapting the later arriving phases before stacking. As expected, this results in all SEWs having the pulse width of the starting S wave GEW. By adapting the later arriving phases to the direct S wave SEW, the onsets are generally sharpened, and the standard deviation is reduced near the wave's onset, as well as over most of the first period of the wave. We thus retain travel times for all phases using the S wave SEW due to its increased stability and well-defined wave onset.

2.4 Results: Travel Times and Other Measurements

2.4.1 Number of Measurements

For the S, SS, SSS, Sdiff, ScS and ScSScS phases we have processed and viewed, over 1.4 million unique seismograms were visually inspected (via the algorithm generated PDF files) from the 360 events. Of these, over 250,000 high quality travel times were retained. The number of retained travel time picks for each phase are given in Table 2.2, along with the number of records viewed and the percentage of viewed records that are retained. The direct S wave by far has the highest rate of measurement success

(56.8%), followed by Sdiff (19.4%) and SS (17.0%). The most viewed phase was ScSScS (with 393,947 records viewed), since it was investigated from 0° in distance to 160°. However, ScSScS only returned confident measurements ~2.6% of the time. The main reasons for records being discarded were low SNR and interference with other phases. The percentage of records measured depends on a number of factors, including event size and radiation pattern combined with event location, the latter which may or may not have stations available at the accepted distance range of the phase of interest. Ridge events in the southern hemisphere were typically noisier than average but used for coverage purposes. Strong deep focus events commonly had more successful picks than average. The direct S wave measurement success varied from less than 1% (for a very noisy southern hemisphere ridge event) to ~89% (for a particularly impulsive and clean South American subduction zone event).

2.4.2 Basic Dataset Attributes

We present some basic attributes of the measurements for each of the six phases in Figure 2.12. The first column displays travel time histograms for the entire dataset. Here we see that the multiple S waves (especially SSS) have a greater spread in the travel time anomalies, as would be expected for a wave with proportionally more of its path in the heterogeneous uppermost mantle. The second column presents histograms of the cross-correlation coefficients (CCCs) between observations and the best-fit SEW. Immediately apparent is that the direct S wave has very high CCCs without much spread. Phases with the longest paths have a greater spread in their CCCs (namely SSS and ScSScS). The SNR measurements (from the average amplitude method) are shown in the third column and emphasize the direct S wave routinely has the highest measured SNRs.

We also present a factor used in the Gaussian function (4th column) that corresponds to pulse width. A Gaussian function, *G*, can be defined as:

$$G(i) = e^{-\frac{i^2}{2g^2}} \tag{2.5}$$

where *i* is the number of time points (and thus the length of the function in time points), *g* is the Gaussian factor (which corresponds to the standard deviation of the function), and *e* is Euler's number. The Gaussian factor histograms display the clearest evidence for the broadening of data pulses for the longer multi-bounce phases (e.g., see SSS compared to SS compared to S, and ScSScS compared to ScS).

2.4.3 Empirical Comprehensive Weight

We have computed an empirical weight measurement for all measured data (for all six seismic phases) that incorporates the SNR (average amplitude method), the CCC between the observation and the best fitting SEW, the misfit of the signal to the best fitting SEW over the first precursory one-period length, the misfit of the signal over the main phase, and the misfit of the signal over the first postcursory one-period length. While empirical, we have found that data with SNR greater than 5 are the best quality, and typically diminish in quality for lower SNR values down to 2, below which we set at a constant value of 0.5. In a similar fashion, the CCC's between observations and the best-fit SEW above 0.92 are common for the good data. This CCC is computed over one period of the main phase (as in Figure 2.8). For this and other measures of quality we present weight functions in Figure 2.14. In addition to SNR and CCC, we show a weight function for the Misfit, measured over a one-period long precursory window ("Misfit Pre I" in Figure 2.8) and a one-period long postcursory window ("Misfit Post I" in Figure 2.8), which documents both precursory and postcursory energy. Misfit is also measured

over one-period of the main phase ("Misfit Main Phase" in Figure 2.8). The weighting values are chosen to emphasize the best measurements based on an inspection across the dataset during data viewing, as well as looking at the trends in Figure 2.12. No weight value drops to zero for our data because all data are human viewed, and only chosen if the data were good quality. We define a comprehensive weight for every travel time measurement from a product of the individual computed weights for each record:

$$w_{comprehensive} = w_{SNR} \times w_{CCC} \times w_{misfit\_pre\_I} \times w_{misfit\_post\_I} \times w_{misfit\_main\_phase} \quad (2.6)$$

where these weights directly correspond to the panels in Figure 2.13. This comprehensive weight is presented with the travel times in the tables that can be downloaded from this work. This comprehensive weight number can be used in modeling experiments for simple weighting approaches, or the individual weights or raw SNR, CCC, and misfits can be used, if preferred.

2.4.4 Full Dataset Attributes

In this study we have measured over two decades of data, built travel time datasets of six dominant SH waves, and measured many attributes of the data. The results are available electronically as a supplement to this study. Table 2.3 lists all of the assembled information. In addition to the travel time anomalies, many data attributes that relate to pulse width (stretching factor, t^*, Gaussian factor), quality (SNR, CCC, misfit), and pulse polarity, are given along with event and station information. This information can be utilized in either forward or inverse modeling experiments or used to retrieve the exact parts of waveforms from the data containing phases of interest. We also estimate the period of every measured phase from displacement recordings. However, there is

some scatter in the estimations due to the presence of longer period energy in displacement components. To help document possible poor period estimates, we compute a best-fit line to our velocity pulse width measurements (i.e., the difference between the pulse end and start, entries 20 and 19, respectively, in Table 2.3) plotted against our estimates of period on displacement recordings. One, two and three standard deviations relative to this best-fit line are computed. We note if the period estimate is an outlier or not by presenting the standard deviation level in Table 2.3 (entry 39).

2.4.5 Geographical Sampling Coverage

Geographical coverage and sampling density are important, since they directly relate to model resolution in both forward and inverse approaches aimed at structure determination. However, due to the geographically restricted earthquake-station geometries, the mantle is unevenly sampled. Most of the earthquakes in our dataset are located on plate boundaries, especially around the circum-Pacific; a majority of the seismic stations are located on continents in the northern hemisphere, including North America, Asia and Europe. Therefore, the northern hemisphere is much better sampled than the southern hemisphere. Our event ranking algorithm has helped to bolster path coverage in less sampled event-station corridors, but uneven sampling persists. Figure 2.14 presents raypath coverage maps of the upper mantle (0-660 km depth) and lowermost mantle (deepest 300 km of the mantle, i.e., the D" layer). Coverage is shown for the individual phases as well as all the phases together. As expected, the northern hemisphere is better sampled than the southern hemisphere, but the southern hemisphere, when all the phases are combined, is sampled everywhere. For the upper mantle, the greatest sampling density generally occurs close to earthquake locations. However, SS,

SSS, and ScSScS provide additional sampling at their surface reflection locations (e.g., like in the central part of both the Pacific and Atlantic Oceans). Regarding the deepest mantle, there has been abundant attention to large low shear velocity provinces, or "LLSVPs" [e.g., *Garnero and McNamara*, 2008; *Lekic et al.*, 2012; *Davies et al.*, 2015a; 2015b; *Cottaar and Lekic*, 2016; *Garnero et al.*, 2016]. We anticipate that these data can help resolution in tomographic studies aiming to sharpen imaging of LLSVP structure.

2.4.6 Stretched versus unstretched EW travel time measurements

The onset time measurements reported in this paper involve stretching each event's SEW to best fit every seismic phase of interest for that event. Here we explore how different these measurements would be if there were no stretching of the empirical wavelet to fit each record. As described in Section 2.3.2 (*Step 5*), a Gaussian function is fit to the stretched SEW for onset time determination of each observation. We also retain the Gaussian function that best fits the SEW that has not be stretched to fit each record (i.e., the representative empirical wavelet of the S wave for each event). We use these two Gaussians to estimate what the onset time difference would be for every measurement in our data set, if we did not stretch the empirical wavelet to fit each record. The stretched minus unstretched wavelet results are plotted in Figure 2.15. A negative (or positive) number in this plot corresponds to the observation being broader (or more narrow, respectively) than the unstretched empirical wavelet. The first order result is that the observed waveform width variability results in significant onset time variability. It is also apparent that longer path lengths (e.g., ScS2, S3) are broadened significantly beyond the unstretched average wavelet of S waves; this is expected from attenuation. However, the spread in the distribution for each phase indicates that using a single pulse for timing

determination (e.g., from an empirical wavelet or synthetic seismograms) will have timing biases due to the variable pulse widths in the observations. We note that there is skew in the distribution (e.g., the direct S wave): slightly more S waves have later onset times with the stretched empirical wavelet compared to an unstretched wavelet. This is due to more S waves being slightly narrower than the mean S wave width, which is caused by exceptionally broadened S waves biasing the mean towards a broader average. This is also indicated in the skewness of the Gaussian factor in Figure 2.12d for S waves. Thus the zero in Figure 2.15 is relative to a mean which is biased by the degree to which events are especially broadened pulse-widths.

2.4.7 Comparison with previous datasets

Various research groups and seismic network operators have implemented largely independent approaches and algorithms to measure travel times. In order to better understand how our new measurements compare with existing datasets, we present relative frequency scatter plots in Figure 2.16, along with a best-fit linear regression between the datasets. For these plots, the travel time anomalies are relative to predictions based on PREM. The datasets contain different earthquakes and stations and therefore we use a summary ray method to find ray paths which are consistent between the datasets. The summary rays average events within 2˚ latitude by 2˚ longitude by 50 km depth bins, and station locations are held constant to within 0.01˚ by 0.01˚. For the relative frequency scatter plots, we calculate the relative occurrence of travel time anomalies between the two datasets in 0.5 x 0.5 sec bins. Bins with a frequency of less than 0.02% are removed to reduce the effect of outliers on the linear regression and weights for each point for the regression is taken as 1/frequency.

The International Seismological Centre (ISC) compiles and publishes global seismic data. One of their primary products is manually determined travel times distributed through the ISC Bulletin. From this database, we extracted ~635,000 direct S arrival travel times between 2000 and 2015 from http://www.isc.ac.uk/iscbulletin/search/arrivals. These times were then reference to PREM using the TauP toolkit (Crotwell et al., 1999) to compute residuals to compare to our measured residuals. The relationship between S times from our EW method and the ISC times is close to linear (Figure 2.16a), but there is significant scatter in the ISC dataset indicated by the low frequency points at the extremes of the *x*-axis. However, the linear regression line has a low offset (0.27) and a slope of 1.13. The correlation coefficient is fairly low (0.51), but the highest probability points (indicated in green to blue) are well aligned with the 1:1 line. This indicates most of our measurements are well aligned with the ISC dataset, but we have fewer large amplitude outliers.

The Array Network Facility (ANF), an Earthscope-USArray funded project largely responsible for the operation of the USArray network, provides manually determined arrival times for the USArray [*Astiz et al., 2014*]. Of the ~2,000,000 total measurements they have published, ~200,000 are direct S arrivals. This dataset has recently been used for tomography in the US [*Golos et al., 2018*] and their results are largely consistent with earlier velocity models for the region. A qualitative comparison of our S times with the ANF times (Figure 2.16b) shows a clear linear relationship with some scatter. The regression line, using 674 bins, has a slope of 1.09 and an offset of 0.47 seconds with a correlation coefficient of 0.87. These values and the scatter plot suggest the datasets are highly consistent.

We also compare our S times against a compilation of times from three mantle tomography studies that measured times by various methods [*Gu et al., 2005; Houser et al., 2008; Ritsema et al., 2011*] in Figure 2.16c. For this comparison, we have applied crustal corrections [*Laske et al., 2013*] so that all times are equally adjusted. The apparent scatter for this comparison is greater than both that of the ANF and the ISC. Also, the regression fit is offset by 1.81 seconds and the slope is well below 1. The correlation coefficient of 0.56 and broad scatter of the data points indicate less agreement in our times with some presented in these past studies. One possible source of the misfit is that determination of travel times through correlation approaches (e.g., between synthetic seismograms and observations) do not consider variable waveform pulse widths in the data which result in systematic biases in reported times. We note that with or without crustal corrections, the nature of the scatter in the comparison in Figure 2.16c does not change.

2.5 Implications and Conclusions

This SEW approach presented in this paper was devised to objectively extract arrival time information from data, while retaining information about wave-shape broadness (from the stretching, t^*, and Gaussian factor information), complexity (from the misfit and cross-correlation measurements), and signal strength (from SNR measurements). Using the same method on all six phases ensures consistency with this global dataset.

All data were visually inspected for every earthquake, but not all data were retained in the dataset. Many data were rejected if significant precursors were present that affected the performance of the algorithm to correctly identify the onset time of the

waveform. The most common example was for SS waves – a significant opposite polarity precursor was commonly present (even when Hilbert-transformed back into the same phase as the direct S wave). While such data may have good SNR, they were omitted from our measurements due to uncertainties in identifying the onset of the phase.

Our initial processing step involved implementation of a bandpass filter between 16 and 100 sec. A shorter period (higher frequency) for the 16 sec upper corner in the filter would yield additional information in waveform broadening as well as the potential for sharper waveform onsets. However, for later arriving phases (as well as all phases for shallow ridge events), noise energy around 10 sec period was commonly present which obscured clear onsets for many data. The 16 sec level for the upper corner of the filter netted significantly more data than broader band filters, yet still retained enough short period information for broadening to be detected and measured. While longer period upper filter corners would likely yield even more measurements, it comes at the cost of losing some shorter period variability in onset times that relates to heterogeneity at smaller scales. As a travel time product oriented paper, this paper is similar to past papers, for example, like Bolton and Masters [2001] who analyzed ~41,000 S waves, or Woodward and Masters [1991], who measured ~6,000 SS waves (referenced to S). The main difference is our larger data set, shorter periods used, and adaptive waveshape fitting.

A powerful aspect of this dataset is the waveform broadening information (e.g., the Gaussian factor). While the computed SEW is a weighted average shape and pulse width, pulse width anomalies relative to the average can be explored in, for example, multi-pathing or attenuation studies (as well as source studies). Our measured travel

times are appropriate for ray theoretical approaches, and thus can be used in travel time tomography and forward modeling analyses. However, many approaches are based on the actual seismic waveforms (instead of onset times). There are different ways the supplementary information files can be utilized for such studies, including: (a) the timing and period information can be used to accurately retrieve the phase of interest from original waveforms, and our data quality information can be utilized to prioritize a data collection scheme; and (b) the Gaussian factor information can be used to construct Gaussian functions which approximate every measured waveform – this can even be used in approaches that rely on synthetic seismograms for correlation with waveforms. These possibilities will depend on the focus of any study to be pursued; we present these possibilities to emphasize the flexibility of the dataset for mapping mantle velocity and/or attenuation heterogeneity. Future analyses might consider fitting multiple Gaussian functions to complex records to better articulate waveform complexities [e.g., Conder, 2015].

We note that event locations used in this study were taken directly from the ISC, and that source relocation has not been done; nor has the quality of source locations been assessed. Work using these data for structure determination (e.g., tomography) should, of course, consider event relocation, as well as elevation and crustal corrections. Here, we focus on an objective measure of raw travel times relative to a reference model, without correcting for ellipticity. While we emphasize the utility of the empirical wavelet construction method for the purpose of travel times and whole earth structure studies, it can also be useful in documenting source and site effects, since the provided information can be parsed in terms of source geometry (e.g., azimuth, depth, take-off angle).

Our method can also be used with other phases, e.g., P waves such as P, PcP, Pdiff, PKP, etc., SV waves such as SKS, SKKS, etc., as well as depth phases (sS, sSS, etc.) and converted waves (e.g., SP, ScP, etc.). Here we chose to develop the method solely with SH waves. Our future plans include application to these other wave types, including higher multiple bounce waves.

We developed an empirical wavelet construction algorithm that adapts an average S wave shape to the six phases S, SS, SSS, Sdiff, ScS, and ScSScS for a collection of 360 global earthquakes. From over 1.4 million viewed seismograms, over 251,000 travel times were made with important waveform and waveshape information retained. Our event selection involved strategies to bolster southern hemisphere coverage, especially in the deep mantle. While coverage still remains denser in the northern hemisphere, the southern hemisphere is fairly well sampled in the deepest several 100 km of the mantle. This was mostly accomplished by adding Sdiff, but also ScS2 and long distance D″ traversing S waves. We have demonstrated that adapting the empirical wavelet to each record can give onset time measurements that differ by up to several seconds from correlative schemes that use a fixed width reference pulse. This can explain some of the differences between our measurements and past studies. All of these measurements and data attributes are publicly available.

2.6 References

Astiz, L., Eakins, J. A., Martynov, V. G., Cox, T. A., Tytell, J., Reyes, J. C., Newman, R. L., Karasu, G. H., Mulder, T., White, M., Davis, G. A., Busby, R. W., Hafner, K., Meyer, J. C., and Vernon, F. L. (2014), The Array Network Facility Seismic Bulletin: Products and an Unbiased View of United States Seismicity, Seismo. Res. Lett., 85 (3), doi:10.1785/0220130141.

Bolton, H., and G. Masters (2001), Travel times of P and S from the global digital seismic networks: Implications for the relative variation of P and S velocity in the mantle, *J. Geophys. Res., 106*(B7), 13527–13540, doi:10.1029/2000JB900378.

Butler, R. (1979), Shear-wave travel times from SS, *Bulletin of the Seismological Society of America*, *69*(6), 1715–1732, doi:10.1121/1.1906594.

Choy, G. L., and P. G. Richards (1975), Pulse distortion and Hilbert transformation in multiply reflected and refracted body waves, *Bulletin of the Seismological Society of America*, *65*(1), 55–70, doi:10.1121/1.1906594.

Conder, J. A. (2015), Fitting Multiple Bell Curves Stably and Accurately to a Time Series as Applied to Hubbert Cycles or Other Phenomena, *Mathematical Geosciences*, *47*(6), 663–678, doi:10.1007/s11004-014-9557-7.

Cottaar, S., and V. Lekic (2016), Morphology of seismically slow lower-mantle structures, *Geophysical Journal International*, *207*(2), 1122–1136, doi:10.1093/gji/ggw324.

Crotwell, H. P., T. J. Owens, and J. Ritsema (1999), The TauP Toolkit: Flexible seismic travel-time and ray-path utilities, *Seismological Research Letters 70*, 154–160.

Davies, C., M. Pozzo, D. Gubbins, and D. Alfè (2015a), Constraints from material properties on the dynamics and evolution of Earth's core, *Nature Geosci*, *8*(9), 678–685, doi:10.1038/ngeo2492.

Davies, D. R., S. Goes, and H. C. P. Lau (2015b), Thermally Dominated Deep Mantle LLSVPs: A Review, in *The Earth's Heterogeneous Mantle*, pp. 441–477, Springer International Publishing, Cham.

Dziewonski, A. M., and D. L. Anderson (1981), Preliminary reference Earth model, Phys. Earth Planet. Inter., 25, 297–356, doi:10.1016/0031-9201(81)90046-7.

Dziewonski, A. M., T. A. Chou, and J. H. Woodhouse (1981), Determination of earthquake source parameters from waveform data for studies of global and regional seismicity, *Journal of Geophysical Research: Solid Earth (1978–2012)*, *86*(B4), 2825–2852, doi:10.1029/JB086iB04p02825.

Ekström, G., M. Nettles, and A. M. Dziewonski (2012), The global CMT project 2004–2010: Centroid-moment tensors for 13,017 earthquakes, *Physics of the Earth and Planetary Interiors*, *200-201*, 1–9, doi:10.1016/j.pepi.2012.04.002.

Futterman, W. I. (1962), Dispersive body waves, *Journal of Geophysical Research: Solid Earth (1978–2012)*, *67*(13), 5279–5291, doi:10.1029/JZ067i013p05279.

Garnero, E. J., A. K. McNamara, and S.-H. Shim (2016), Continent-sized anomalous zones with low seismic velocity at the base of Earth's mantle, *Nature Geosci*, *9*(7), 481–489, doi:10.1038/ngeo2733.

Garnero, E. J., and A. K. McNamara (2008), Structure and Dynamics of Earth's Lower Mantle, *Science*, *320*(5876), 626–628, doi:10.1126/science.1148028.

Golos, M., Fang, H., Yao, H., Zhang, H., Burdick, S., Vernon, F., Schaeffer, A., Lebedev, S., and van der Hilst, R. D. (2018), Shear Wave Tomography Beneath the United States Using a Joint Inversion of Surface and Body Waves, J. Geophys. Res., Solid Earth, doi:10.1029/2017JB014894.

Gu, Y. H., A. L. Lerner-Lam, A. M. Dziewonski, and G. Ekström (2005), Deep structure and seismic anisotropy beneath the East Pacific Rise, Earth Planet. Sci. Lett., 232, 259–272, doi:10.1016/j.epsl.2005.01.019.

Houser, C., G. Masters, P. Shearer, and G. Laske (2008), Shear and compressional velocity models of the mantle from cluster analysis of long-period waveforms, *Geophysical Journal International*, *174*(1), 195–212, doi:10.1111/j.1365-246X.2008.03763.x.

Inoue, H., Y. Fukao, K. Tanabe, and Y. Ogata (1990), Whole mantle P-wave travel time tomography, *Physics of the Earth and Planetary Interiors*, *59*(4), 294–328, doi:10.1016/0031-9201(90)90236-Q.

Kennett, B. L. N., E. R. Engdahl, and R. Buland (1995), Constraints on seismic velocities in the Earth from traveltimes, *Geophysical Journal International*, *122*(1), 108–124, doi:10.1111/j.1365-246X.1995.tb03540.x.

Kustowski, B., G. Ekström, and A. M. Dziewonski (2008), Anisotropic shear-wave velocity structure of the Earth's mantle: A global model, *Journal of Geophysical Research: Solid Earth (1978–2012)*, *113*(B6), B06306, doi:10.1029/2007JB005169.

Laske, G., G. Masters, Z. Ma, and M. Pasyanos (2013), Update on CRUST1.0—A 1-degree Global Model of Earth's Crust, Geophys. Res. Abstr.,15, EGU2013-2658.

Lekic, V., S. Cottaar, A. Dziewonski, and B. Romanowicz (2012), Cluster analysis of global lower mantle tomography: A new class of structure and implications for chemical heterogeneity, *Earth and Planetary Science Letters*, *357-358*, 68–77, doi:10.1016/j.epsl.2012.09.014.

Lou, X., and S. van der Lee (2014), Observed and predicted North American teleseismic delay times, *Earth and Planetary Science Letters*, *402*, 6–15, doi:10.1016/j.epsl.2013.11.056.

Lou, X., Van Der Lee, S., and Lloyd, S. (2013), AIMBAT: A python/matplotlib tool for measuring teleseismic arrival times, *Seismological Research Letters, 84*(1), 85-93.

Ni, S., and Helmberger, D. V. (2003), Ridge-like lower mantle structure beneath South Africa, *J. Geophys. Res., 108*, 2094, doi:10.1029/2001JB001545.

Pavlis, G. L. and Vernon, F. L. (2010), Array processing of teleseismic body waves with the USArray, *Computers and Geosciences, 36*(7), 910–920.

Ritsema, J., A. Deuss, H. J. Van Heijst, and J. H. Woodhouse (2011), S40RTS: a degree-40 shear-velocity model for the mantle from new Rayleigh wave dispersion, teleseismic traveltime and normal-mode splitting function measurements, *Geophysical Journal International*, *184*(3), 1223–1236, doi:10.1111/j.1365-246X.2010.04884.x.

Ritsema, J., and H. J. van Heijst (2002), Constraints on the correlation of P- and S-wave velocity heterogeneity in the mantle from P, PP, PPP and PKPab traveltimes, *Geophysical Journal International*, *149*(2), 482–489, doi:10.1046/j.1365-246X.2002.01631.x.

Rost, S. (2002), Array seismology: Methods and applications, *Rev. Geophys.*, *40*(3), 1008, doi:10.1029/2000RG000100.

Schaff, D. P., and F. Waldhauser (2005), Waveform Cross-Correlation-Based Differential Travel-Time Measurements at the Northern California Seismic Network, *Bulletin of the Seismological Society of America*, *95*(6), 2446–2461, doi:10.1785/0120040221.

Simmons, N. A., A. M. Forte, L. Boschi, and S. P. Grand (2010), GyPSuM: A joint tomographic model of mantle density and seismic wave speeds, *J. Geophys. Res.*, *115*(B12), B12310, doi:10.1029/2010JB007631.

To, A., Romanowicz, B., Capdeville, Y., and Takeuchi, N. (2005), 3D effects of sharp boundaries at the borders of the African and Pacific Superplumes: Observation and modeling, *Earth Planet. Sci. Lett., 233*, 137-153, doi:10.1016/j.epsl.2005.01.037.

van der Hilst, R. D., S. Widiyantoro, and E. R. Engdahl (1997), Evidence for deep mantle circulation from global tomography, *Nature*, *386*(6625), 578–584, doi:10.1038/386578a0.

Vandecar, J. C., and R. S. Crosson (1990), Determination of teleseismic relative phase arrival times using multi-channel cross-correlation and least squares, *Bulletin of the Seismological Society of America*, *80*(1), 150–169, doi:10.1029/JB091iB14p13873.

Woodhouse, J. H., and A. M. Dziewonski (1984), Mapping the upper mantle: Three-dimensional modeling of earth structure by inversion of seismic waveforms, *Journal of Geophysical Research: Solid Earth (1978–2012)*, *89*(B7), 5953–5986.

Woodward, R. L., and G. Masters (1991), Global upper mantle structure from long-period differential travel times, J. Geophys. Res., 96(B4), 6351–6377, doi:10.1029/90JB01445.

Table 2.1. Seismic networks used in this study for global dataset construction. Data were collected using software associated with each data center, including SOD (Standing Order for Data), BREQ_FAST (Batch Requests, Fast), NetDC (Networked Data Center Protocol) and AutoDRM (an email-based request tool).

Network name	URL
Incorporated Research Institutions for Seismology (IRIS)	http://www.iris.edu
Observatories & Research Facilities for European Seismology (ORFEUS)	http://www.orfeus-eu.org
Northern California Seismic Network (NECDC)	http://www.ncedc.org/ncsn
F-net Broadband Seismograph Network	http://www.fnet.bosai.go.jp
Canadian National Seismic Network (CNSN)	http://www.earthquakescanada.nrcan.gc.ca/stndon/CNSN-RNSC

Table 2.2. Total number of travel time measurements in this study for all events, along with the total number of records viewed and inspected, and the percent of the viewed data that produced successful travel time measurements.

Phase Name	# accepted	# evaluated	% data measured
S	123,946	218,094	56.8
SS	53,505	314,259	17.0
SSS	11,927	207,429	5.7
Sdiff	28,499	146,561	19.4
ScS	23,758	158,850	14.9
ScSScS	10,303	393,947	2.6
All	251,939	1439,140	17.5

Table 2.3. List of data attributes compiled and computed in this study that are shared in electronic supplemental information. The number in the first column of the table below corresponds to the column number in the archived data file.

#	Information	Description
1	Station name	The 3-5 character station name code
2	Network name	The two-digit code for the seismographic network
3	Distance	Epicentral distance between earthquake and station in degrees
4	Station latitude	Station location latitude in degrees[1]
5	Station longitude	Station location longitude in degrees[1]
6	Event latitude	Earthquake hypocentral location latitude in degrees[2]
7	Event longitude	Earthquake hypocentral location longitude in degrees[2]
8	Event depth	Earthquake hypocentral location depth in km[2]
9	Event magnitude	Earthquake moment magnitude[2]
10	Origin time	Earthquake origin time[2]
11	Azimuth	Azimuth from earthquake to station (in degrees)
12	Back azimuth	Back azimuth measured at station clockwise back to earthquake (in degrees)
13	Phase name	Either S, SS, SSS, Sdiff, ScS, or ScSScS
14	Measured time	Travel time anomaly of phase onset relative to PREM (observed minus PREM)
15	Predicted time	Travel time prediction of the PREM model
16	Amplitude	Amplitude of the peak of the phase, on instrument-deconvolved velocity recordings
17	Radiation pattern	The predicted amplitude between [−1,1] using the SH radiation pattern
18	Flipped polarity flag	Record was modeled after flipping the record's polarity (which was explored for a low radiation pattern amplitude range [−0.15,0.15]), flagged as 0 if radiation pattern prediction is correct, otherwise 1
19	Phase start	The start time, relative to PREM prediction, of the beginning of the time window used to define one pulse width of the phase on velocity recordings, measured at the 10% amplitude level preceding the wave peak (used to auto-define the Misfit measurement windows)
20	Phase end	The end time, relative to the PREM prediction, of the end of the time window used to define one pulse width of the phase on velocity recordings, measured at the 10% amplitude level following the wave peak (used to auto-define the Misfit measurement windows)
21	$SNR_{average\_amp}$	The signal-to-noise measurement from the average amplitude of the signal to the average amplitude of the noise, as in equation (2)

22	$SNR_{peak\text{-}trough}$	The signal-to-noise measurement from the maximum peak-to-trough measurement within one period of the signal compared to noise, as in equation (3)
23	SNR_{max_peak}	The signal-to-noise measurement from the maximum peak in the signal window compared to the maximum peak in the entire noise window, as in equation (4)
24	$Misfit_{SIGNAL}$	The average difference between the phase and the best-fit SEW over one period (see Equation 1)
25	$Misfit_{PRE}$	As above, except over one period preceding the phase
26	$Misfit_{POST}$	As above, except over one period following the phase
27	$Misfit_{PRE\_2T}$	As above, except over one period preceding $Misfit_{PRE}$
28	$Misfit_{POST\_2T}$	As above, except over one period following $Misfit_{POST}$
29	t*	The best-fit *t** value that, when convolved with the SEW, gives the best fit to records that are broader than the SEW
30	Stretch factor	A measure of the amount the SEW has to be narrowed to fit records that are narrower than the SEW
31	$CCC_{[rec,SEW]}$	Cross-correlation coefficient between observed record and the best-fitting SEW adapted to the record
32	$CCC_{[rec,GEW]}$	Cross-correlation coefficient between observed record and GEW, which measures the record's fit to the average phase shape for the event
33	$g_{\_best\text{-}fit\_SEW}$	Gaussian factor of the best-fitting Gaussian function (equation 5) to a record's best-fitting SEW
34	$g_{\_event\_SEW}$	Gaussian factor of the best-fitting Gaussian function (equation 5) to the S wave SEW for the event (i.e., unstretched, un-*t**'ed, as in *Step 3*, Section 2.3.1)
35	$Misfit_g$	The misfit measured between $g_{\_best\text{-}fit\_SEW}$ and $g_{\_event\_SEW}$ (computed as in equation 1) which provides a different measure of record broadening
36	$w_{\_comprehensive}$	An empirical comprehensive weight value for each data (see equation 6)
37	Noise window traffic flag	Records that have "traffic" (other seismic waves) predicted to arrive in the noise window (of the SNR measurement) are flagged as 1, otherwise 0
38	Period	Estimated period of the phase, from the start and end of the pulse measured at 0.1 amplitude (when peak is set to 1) measured on displacement recordings
39	Period Flag	"1" if period measurement is within 1 standard deviation from best-fit line of displacement period and velocity pulse width measurements, "2" if period measurement is in the 1 to 2 standard deviation population, and "3" if in the population greater than standard deviation of 2.

[1]As provided by the data agencies in Table 2.1.
[2]As provided by ISC (International Seismological Center)

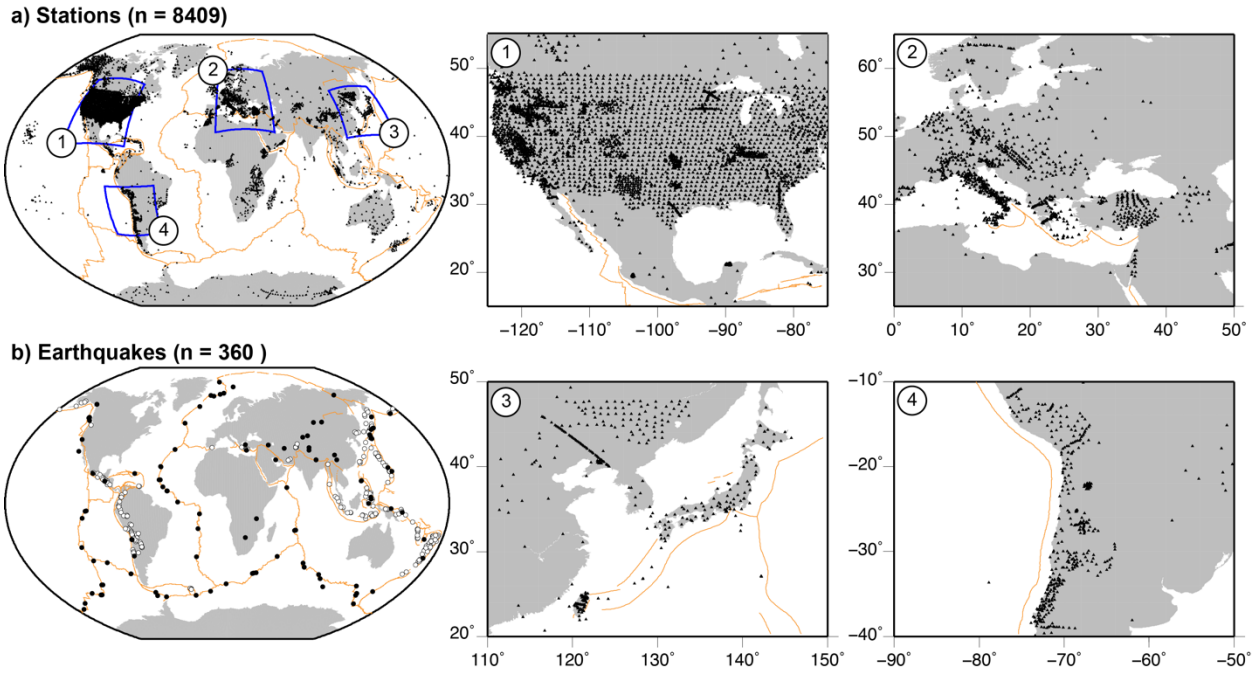

Figure 2.1. Geographical distribution of (a) stations with four regions (blue boxes) shown enlarged on the right, and (b) earthquakes used in this study. The "*n*" value above each map indicates the number of stations (top) or events (bottom). Earthquakes deeper than 50 km are white-filled circles with black outlines; earthquakes shallower than 50 km are solid black circles. Plate boundaries are orange lines. The zoomed in station panels correspond to relatively dense station coverage in the US (panel 1), Europe (panel 2), Japan and west Asia (panel 3), and western South America (panel 4).

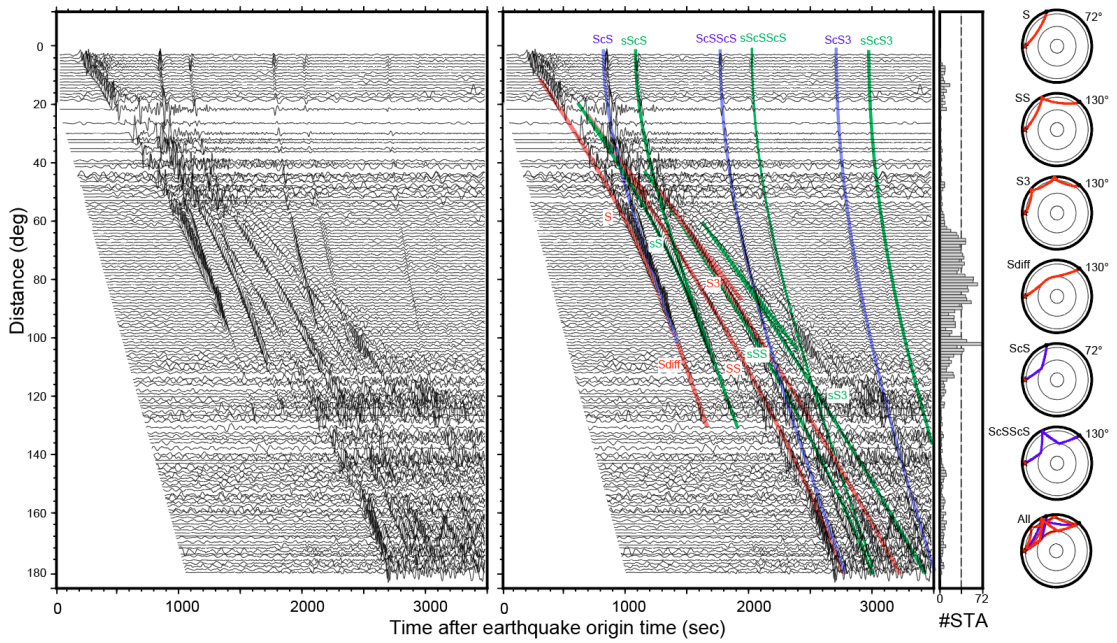

Figure 2.2. Record section for an event on May 28, 2012 (origin time 05:07, source depth 591.1 km). Nearly an hour of record length is shown for the transverse component of motion velocity recordings; each trace is a stack of all records in a 1.5° distance window. The number of records for each stack is plotted in the histogram to the right. The two record sections are identical, except the one on the right has travel time predictions of S, SS, SSS and Sdiff (red lines), the core-reflected ScS, ScSScS, and ScSScSScS waves (blue lines). Green lines correspond to depth phases. The small cross-sections to the right display example ray paths of the 6 main phases used in this study: S, SS, SSS, Sdiff, ScS, and ScSScS. The cross-section marked by "All" combines the paths of the six phases.

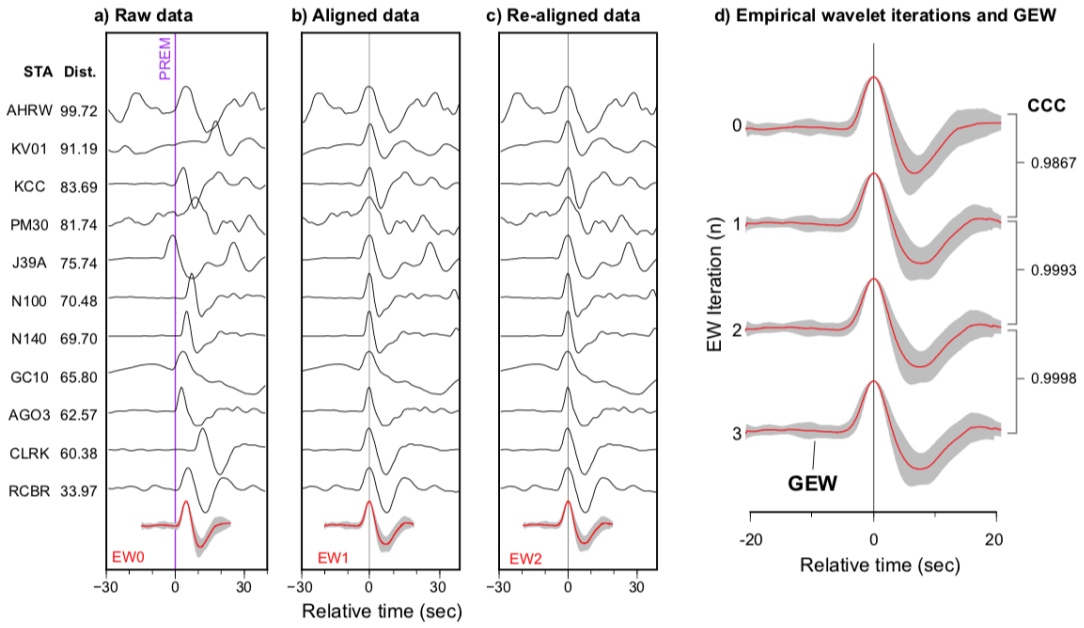

Figure 2.3. Sample seismograms from the May 28, 2012 event of Figure 2 (station names and epicentral distance in degree given on the left). In all panels the traces are centered on the direct S wave, and plotted relative to the S wave time predicted by PREM. (a) S wave recordings displayed in order of epicentral distance, then directly stacked to generate the zeroth stack, EW0, which is shown at the bottom in red with one standard deviation shown as the grey shaded region. (b) The records from (a) have been aligned with EW0 using cross-correlation, and then restacked to make EW1, shown at the bottom. (c) Same as in (b), except the records are now aligned with EW1 and restacked to make an update, EW2, shown at the bottom. This process is repeated until the stack does not update any longer. (d) The empirical wavelet iterations are shown, from EW0 to EW3 (red lines) with one standard deviation (gray shading). The cross-correlation coefficient between successive empirical wavelets is shown on the right. The final iteration (EW3 here) is called the General Empirical Wavelet (GEW).

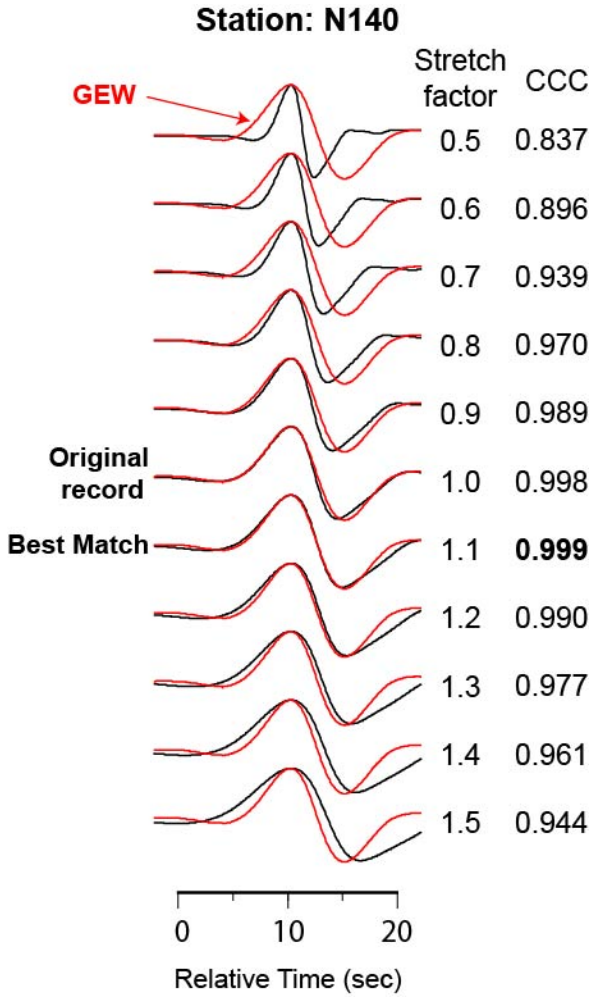

Figure 2.4. Observed S-wave pulse from station N140 (of Figure 3) shown with different degrees of time compression/expansion (black traces, original record has stretch factor of 1), each compared with the GEW from Figure 3d (red traces, which are all identical). The stretched record that gives the highest cross-correlation coefficient (CCC) with the GEW (for this example, the bold black "Best Match" trace with stretch factor 1.1) is retained to make the stretched empirical wavelet (SEW). The stretching factors and corresponding cross-correlation coefficients are listed on the right side.

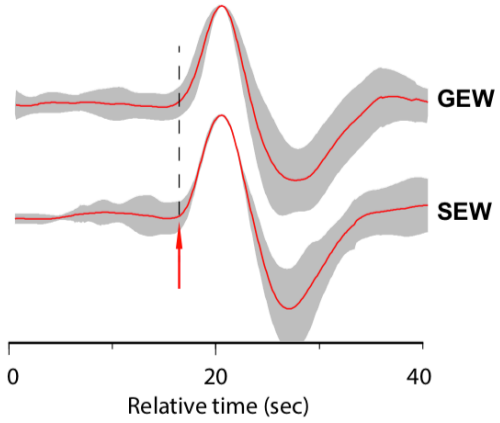

Figure 2.5. Comparison of the General Empirical Wavelet (GEW) and Stretched Empirical Wavelet (SEW) for the records of Figure 3. The GEW and SEW are shown as red traces with gray shading representing one standard deviation around the stack. The arrow indicates the onset in the SEW, which is sharper than the more rounded onset in the GEW due to averaging records of variable width. Also apparent in the SEW is the significantly reduced variability around the main upswing and following downswing of the S wave.

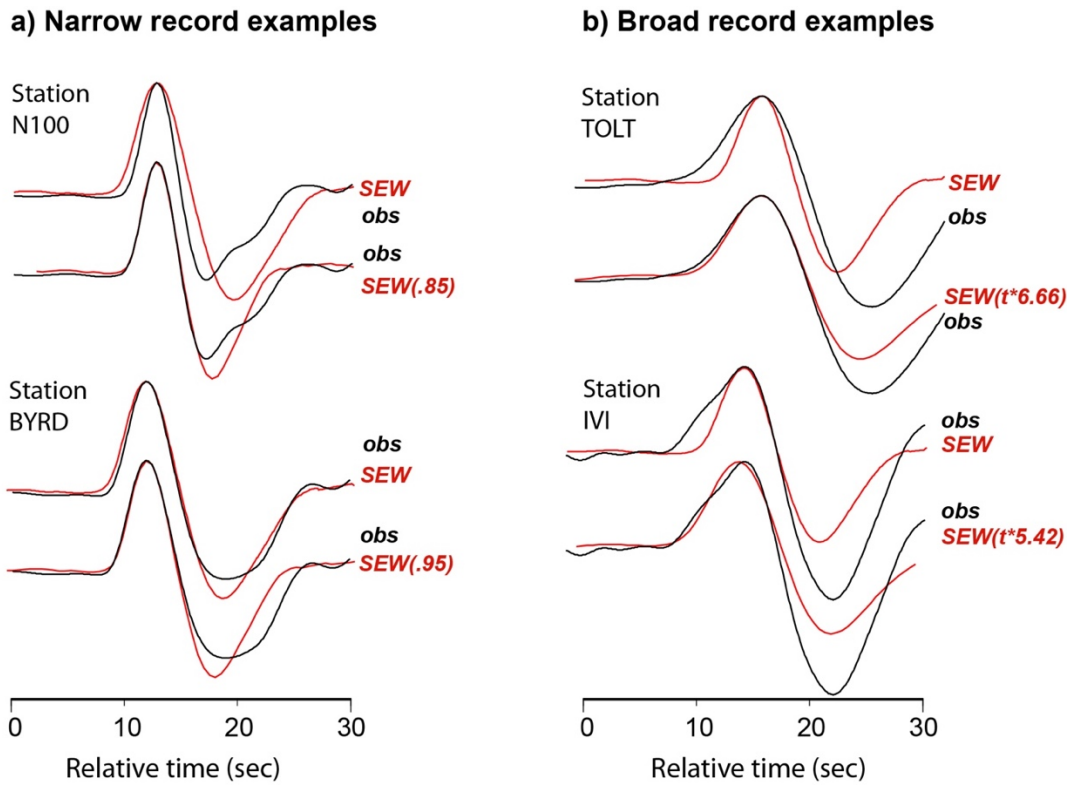

Figure 2.6. Examples showing the SEW being (a) narrowed and (b) broadened to maximize the correlation with observations from the May 28, 2012 event. (a) Observed S waves from stations N100 (top) and BYRD (bottom); black traces are the data, and red traces are the unaltered SEW and the SEW that has been narrowed to best match the observation. The time reduction factor that gives the best fit is shown in parentheses, e.g., for station N100 the best-fit SEW has been narrowed to 0.85 its original time width. (b) as in (a) except the observations are broader than the SEW. For these stations, the SEW is convolved with a t^* operator to obtain a best fit with the observation. For the two examples, the t^* value is given in parentheses. The examples in this figure demonstrate the variable pulse widths in the data, and how perturbing the SEW can result in matching the observed shapes.

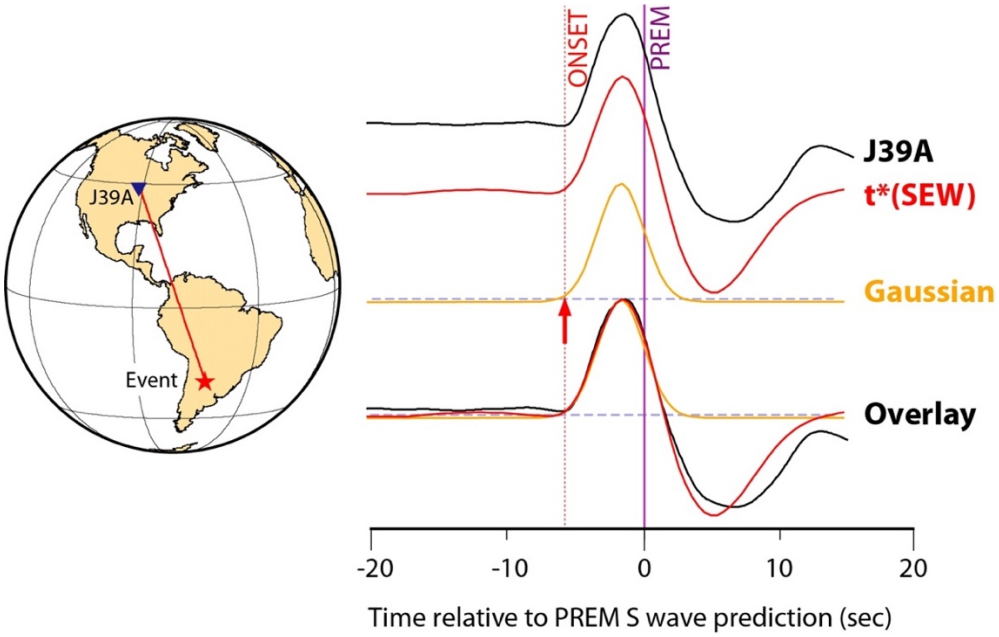

Figure 2.7. Onset time determination from adapting a Gaussian function (orange trace) to match the SEW (red trace) that best fits the S wave observation (black trace) for the deep focus earthquake (red star on globe on left, also, Figure 3) recorded at station J39A (blue triangle in globe). The SEW has been convolved with a t^* operator to best fit the observation. The PREM time for this recording is shown by the purple vertical line. The onset time corresponding to the 1% amplitude level of the Gaussian function is indicated by the red arrow (and vertical red dashed line), which is assigned to the observation. The three traces are overlain in the lower right. This method indicates this record has a travel time anomaly of –6.7 sec relative to PREM.

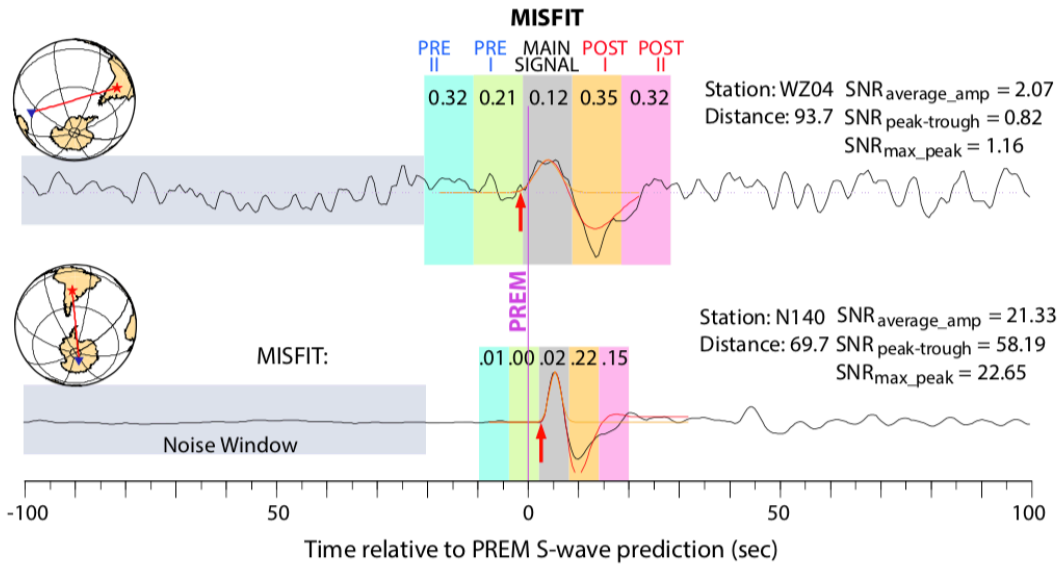

Figure 2.8. Two example S wave records (stations WZ04 and N140) are shown along with misfit and SNR measurement windows and values. These are from the same May 28, 2012 event as in previous figures. The small globes to the left present the event (red star), stations (blue triangles), and great circle paths (red lines). Purple vertical line is the PREM predicted time. Two pre- and post-cursory time windows of 1 pulse width length flank the pulse width of the main signal, where misfit measurements are made (numbers at the top of the colored boxes). The 80 sec gray-shaded window on the left is the noise window used to calculate the different SNR measurements, which are listed to the right. See main text for additional details.

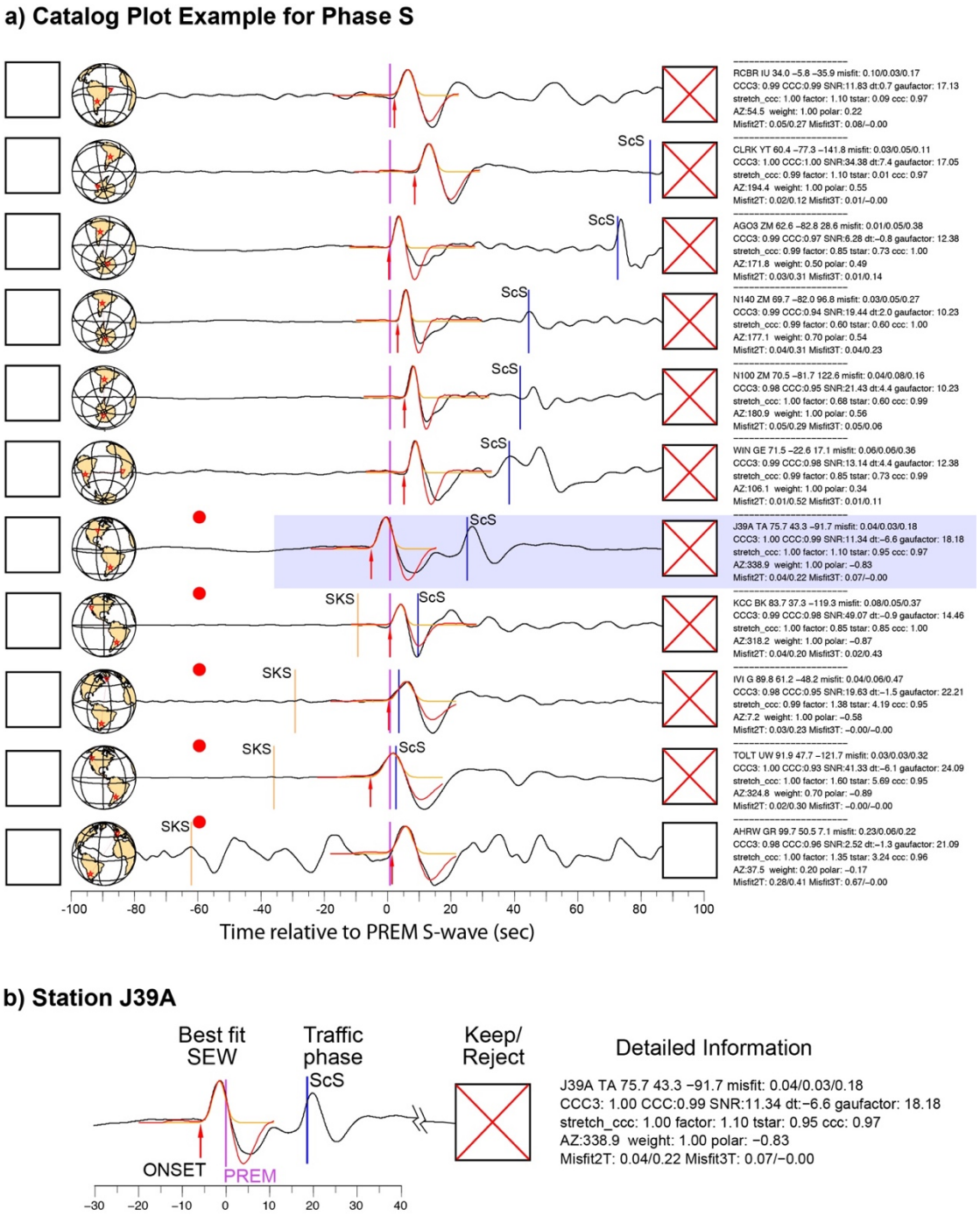

Figure 2.9. Example of portion of catalog PDF page of direct S waves with overlain SEW used for human reviewing. These data are from an event on May 28th, 2012. (a) A portion of the PDF page that shows from left to right: an alphanumerical text input box for the user to add a code or time shift if desired/necessary; a small globe showing the

event (star)-to-station (triangle) geometry; 200 sec of the observed transverse component velocity recording with the S wave near zero time, aligned according to the PREM prediction (purple line at zero); a box with a red “X” (or empty) which logs the algorithm’s decision for the data being good or poor, which the reviewer can update, and a text block with information about the station and all the measurements made. (b) A zoom in of one of the records (station J39A, shaded region in top panel), which shows: the best-fit SEW overlaid on the S-wave (red trace), the best-fit Gaussian (orange trace), travel time predictions of other phases (here, ScS is present), a red dot that indicates the code flipped the polarity so that the phase of interest is a positive pulse in the plot so all polarities are uniform, the onset time determined from the Gaussian (red arrow), the PREM predicted time (purple line), the retain/reject box, and the detailed information block, with detailed measurements such as travel time anomaly, SNR, CCC, misfit, and Gaussian factor.

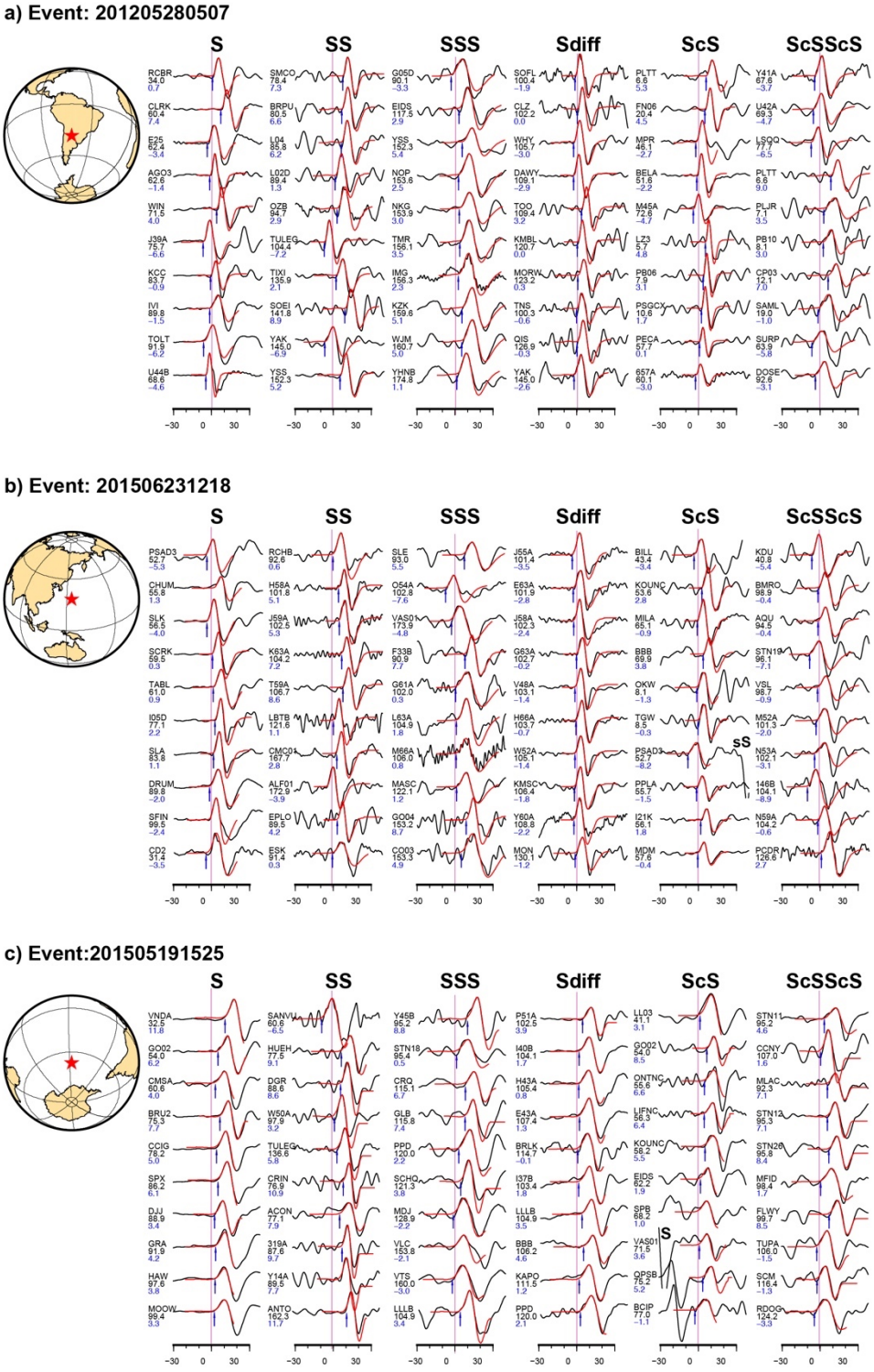

Figure 2.10. Example records of S, SS, SSS (where SS and SSS are put into the phase of S), Sdiff, ScS, and ScSScS for three events. (a) Event 201205280507 (where the title is in

the form yyyymmddhhss, where yyyy = year, mm = month, dd = day, hh = hour, and ss = seconds of the origin time), with latitude, longitude, depth, and moment magnitude of [–28.02°, – 53.11°, 591.1 km, 6.7]. The globe to the left shows the event location (red star). (b) As in (a) except for Event 201506231218, with latitude, longitude, depth, and moment magnitude of [27.74°, 139.72°, 460 km, 6.5]. (c) as in (a), except for Event 201505191525 with latitude, longitude, depth, and moment magnitude of [–54.33°, –132.16°, 7.2 km, 6.7]. Black traces are the raw records displayed relative to the PREM predicted time and arranged according to increasing distance. The red traces correspond to the S wave SEW for each event, adapted to best-fit every arrival, and overlaid with each phase. The blue arrows correspond to the solution onset times, derived from the best-fitting Gaussian (not shown) to each SEW. The station name for each record is listed on the left and underlain by the epicentral distance and travel time anomaly relative to PREM (blue text). Some additional arrivals are present for some records, and the known phases are named.

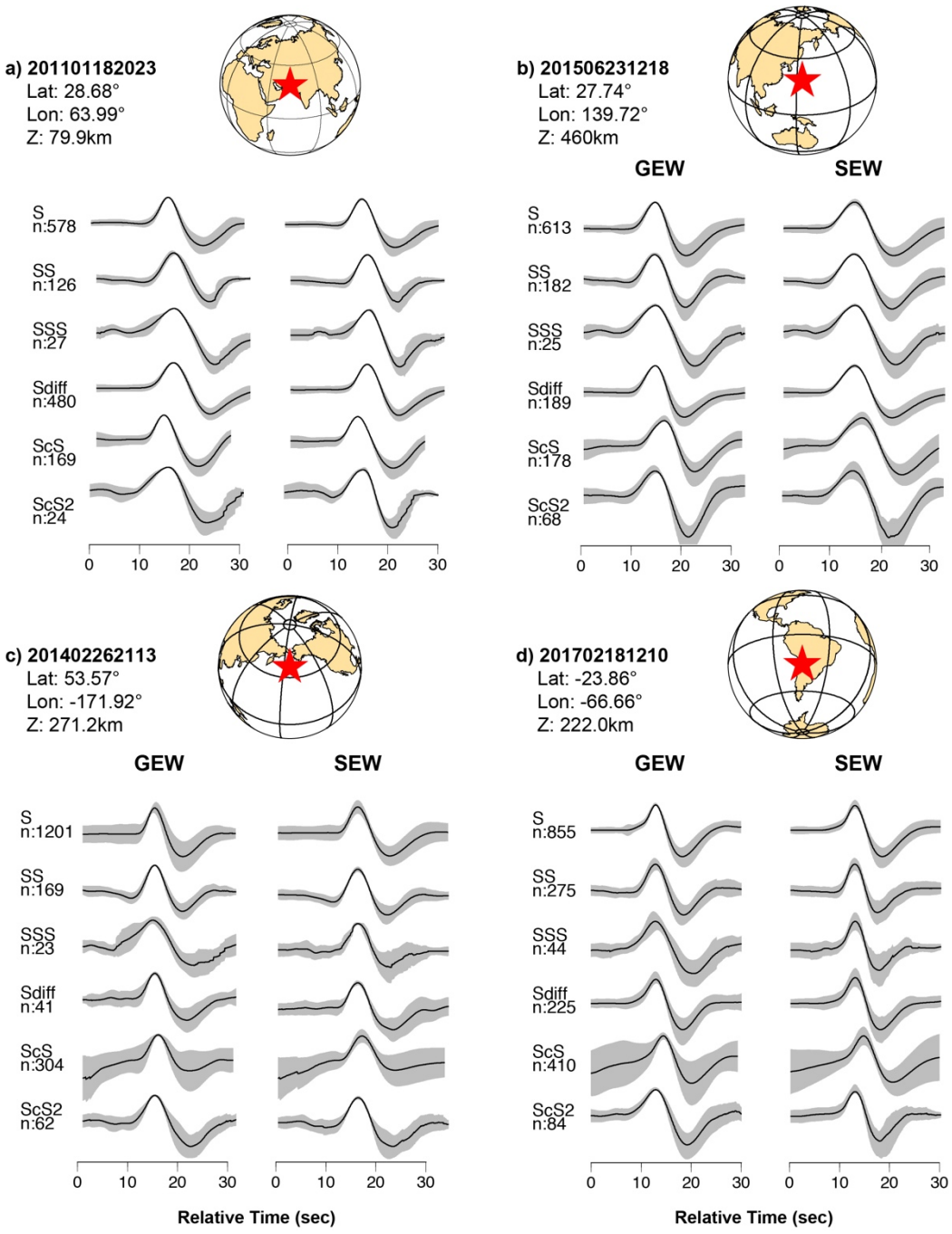

Figure 2.11. Generic Empirical Wavelets (GEW) and Stretched Empirical Wavelets (SEW) with standard deviation for four example events in panels (a) through (d), for the seismic phases S, SS, SSS, Sdiff, ScS, and ScSScS, where SS and SSS have been Hilbert transformed back into the phase of S. Also shown are the event date and origin time (in format YYYYMMDDMMSS, see Figure 10), latitude (Lat), longitude (Lon), and source

depth (Z), and the location of each event (red star) in a small global map in the upper left of each panel. Each panel has two columns: the left column presents the GEW (along with one standard deviation, gray shading) made from stacking that phase (the phase name is noted on the left, along with number of records used in the stack). The right column is the SEW stack of each phase, using the S wave GEW (top trace in left column) as the reference shape to which records of each phase are adapted to (see text for more details).

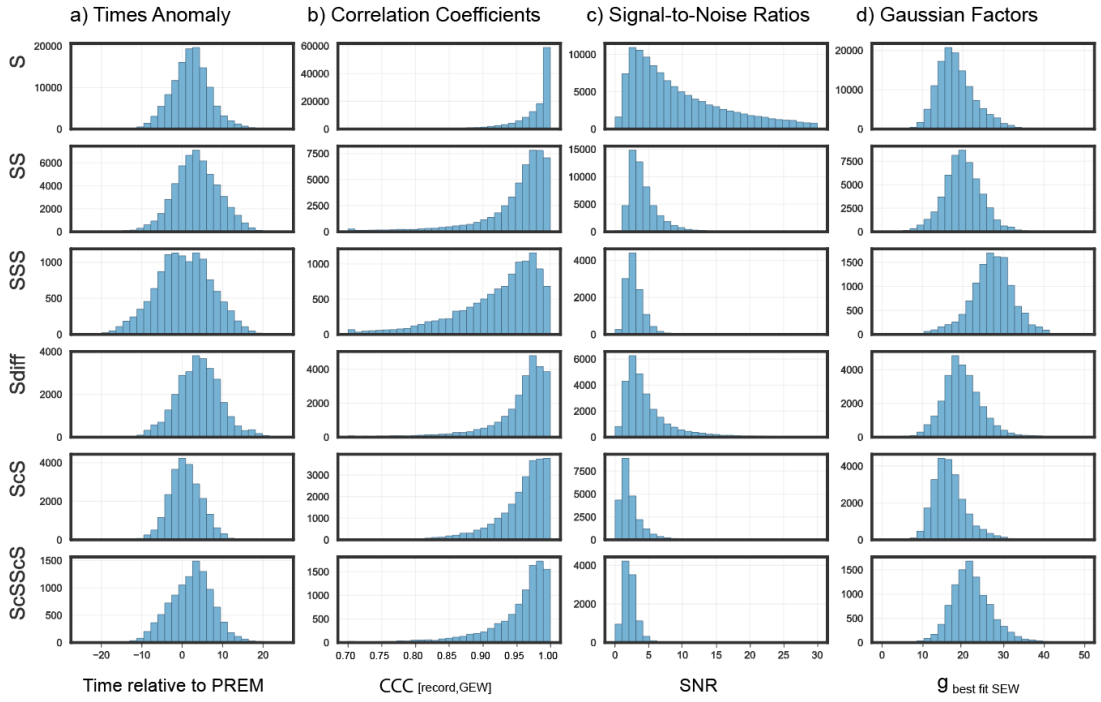

Figure 2.12. For the six phases studied, histograms are shown for (a) travel time anomaly relative to PREM predictions, (b) Cross correlation coefficients between the phase of interest and the best-fitting SEW, (c) the SNR measured from the average amplitude method, and (d) the Gaussian factors for the Gaussian function that best matches the SEW which fits each record.

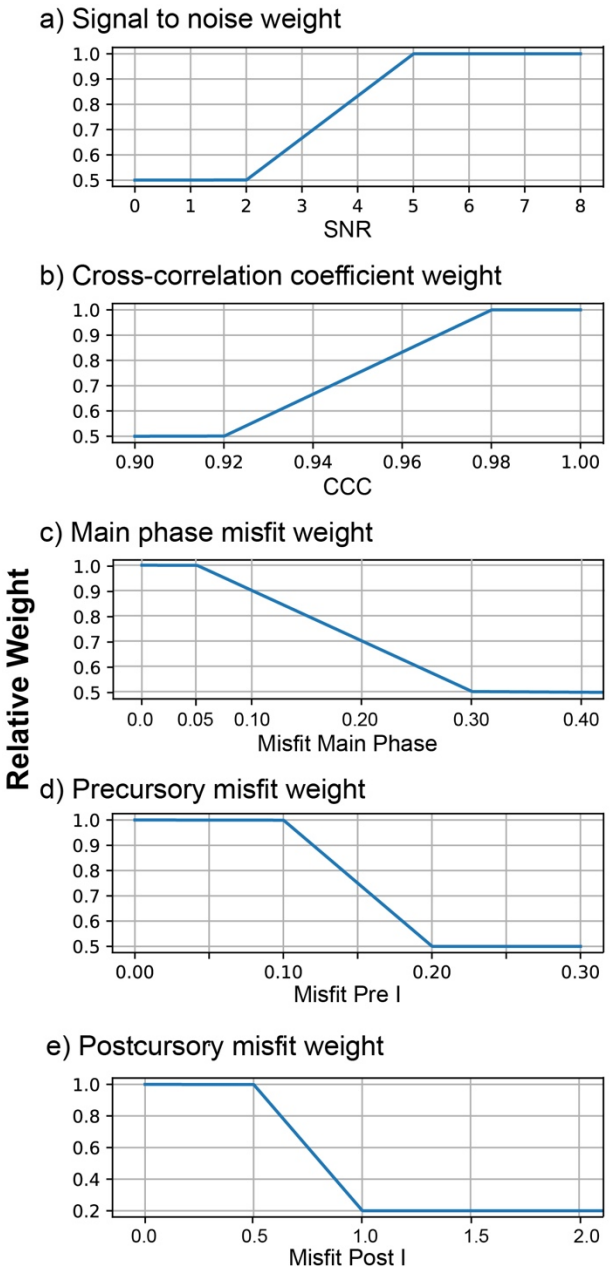

Figure 2.13. Weighting functions for various measurements made in our data analyses: (a) SNR measured by the average amplitude method (equation 2); (b) CCC between the best-fitting SEW and the observation; (c) the Misfit measured over the pulse width of the main phase of interest; (d) Misfit measured over one pulse width before the arrival of the main phase, i.e., the precursory energy window ("Misfit Pre I" in Figure 8); and (e) the Misfit measured over one pulse width after the main phase ("Misfit Post I" in Figure 8).

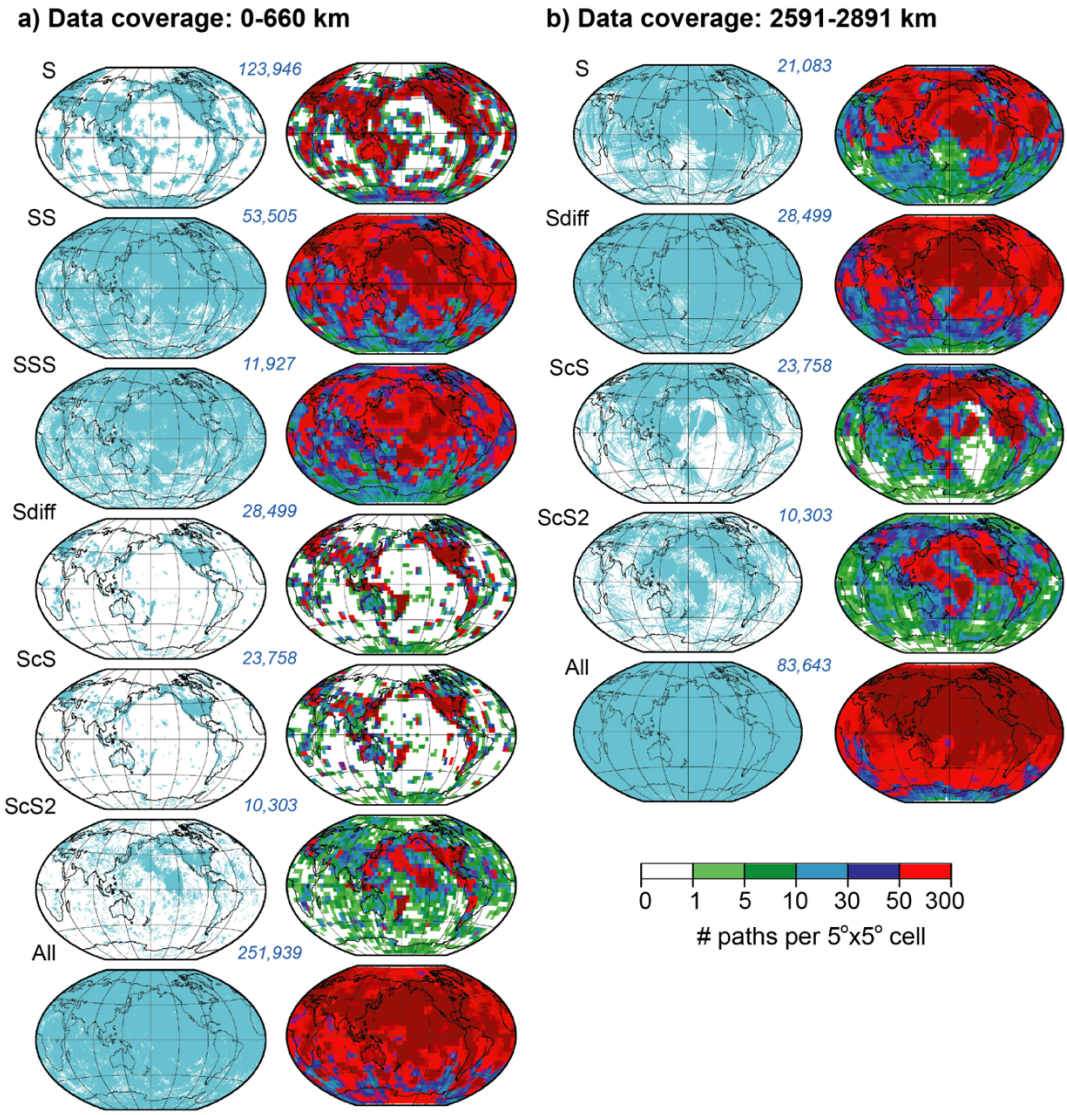

Figure 2.14. (a) Ray path coverage map in the upper mantle (0-660 km) for the six seismic phases studied here; the upper mantle portion of every ray path is drawn as a light blue line (left column). The panel at the bottom named "All" has all ray paths plotted on the same panel. The right column shows the ray path sampling density in 5° by 5° cells. Red corresponds to high sampling density (see scale bar). (b) As in (a) except for the deepest 300 km of the mantle, the D″ layer. Only phases with robust D″ sampling are shown. As with (a), the "All" panel combines the D″-sampling waves.

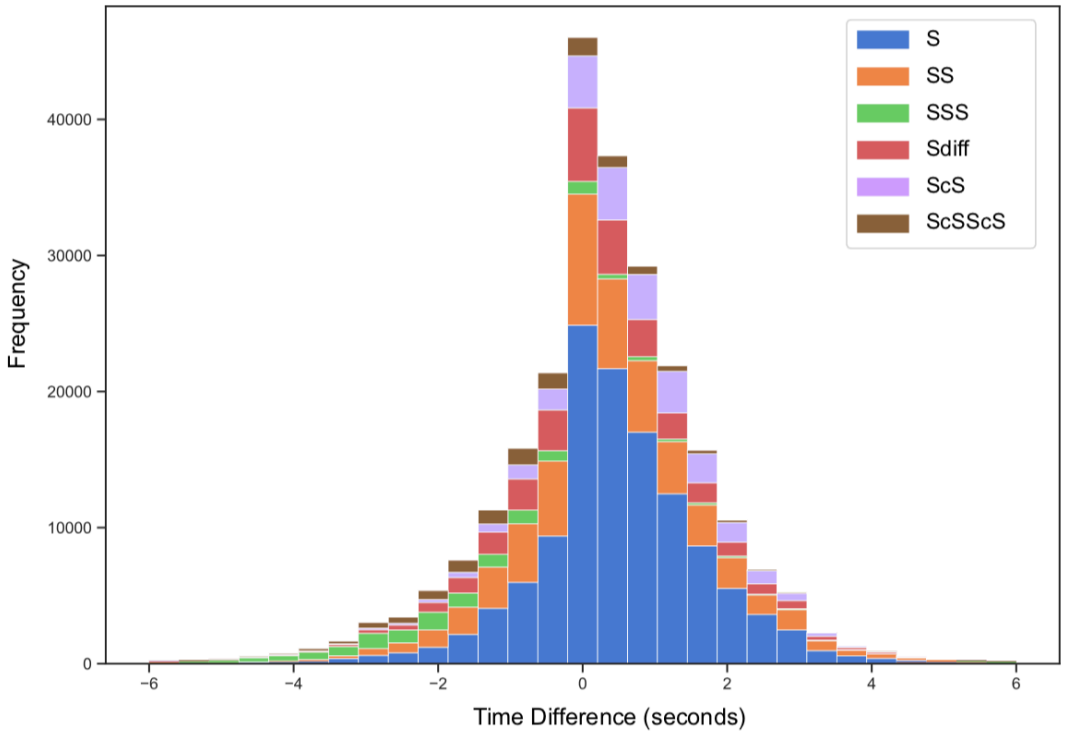

Figure 2.15. Stacked histogram of the difference between the onset time of the Gaussian function of best-fit SEW of each record and the Gaussian function of the average (unstretched) S wave of each event (the first computed SEW, as in Figure 5). The stretched SEW Gaussian minus the unstretched GEW Gaussian is shown. Each color represents histogram from a different phase.

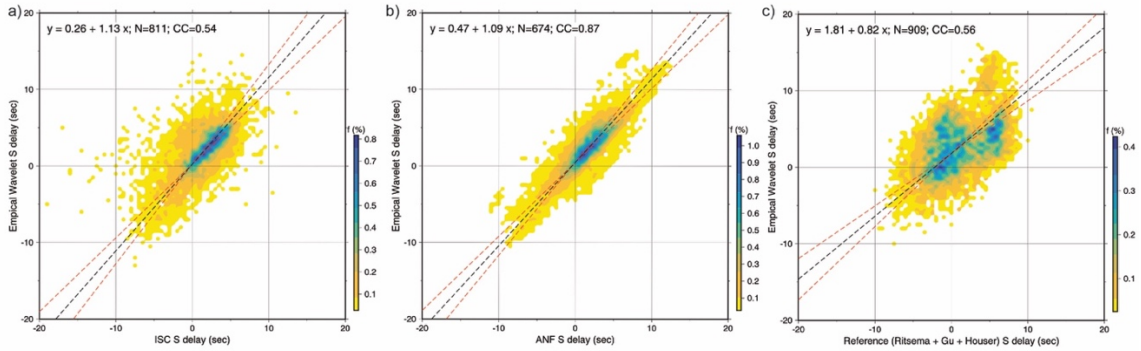

Figure 2.16. Measured travel time anomalies from our empirical wavelet method are compared to measurements from (a) the ISC bulletin, (b) the ANF catalog, and (c) times reported in three tomographic studies. Summary ray times constructed for small event and station bins are compared in the figure (thus, every plotted point represents one summary ray comparison). The equation for the best-fitting line is given (black dashed line), along with uncertainty bounds (red dashed lines) for the best fitting line. Also shown is the number of plotted summary ray comparisons (N) and the correlation coefficient (CC) for each best fitting line. Symbols are colored according to the frequency for every 0.5 by 0.5 sec time cell on the plot. See text for additional details.

CHAPTER 3

3 SUPERVISED CLASSIFICATION ON SEISMIC DATA

Abstract

To assist our manual confirmation process in Empirical Wavelet construction, an empirical classification model is implemented to distinguish high quality and noisy seismic records based on rules associated with signal-to-noise ratio, cross-correlation coefficients and waveform misfit measurements. However, the performance of the model is not effective enough. A better classification model that is able to accurately distinguish high quality data from the noise is highly desired, which combined with the automated Empirical Wavelet construction would allow fully automated seismic data characterization and information retrieval. Supervised machine learning model learns a function by mapping the input to the output labels based on the example input-output pairs [*Russell & Norvig, 2010*]. It has been successfully implemented in many applications that involves large and complex dataset. After exploring the correlation of the comprehensive measurements that we have documented in Empirical Wavelet construction and the target labels (high quality versus low quality), a binary classification model is trained on the Empirical Wavelet dataset. A fast, distributed, high performance gradient boosting framework (LightGBM) is used during the training, additionally boosting and bagging techniques have been implemented in the training process to reduce bias and avoid overfitting. The final model is tested in the reserved dataset, which shows significant improvement comparing to the empirical classification model.

3.1 Introduction

In Chapter 2, we have constructed Empirical Wavelet construction method to automatically document seismic measurements, including travel time and waveform information [*Lai et al.*, 2019]. This method is automated and applied to 360 events and resulted in a high-quality seismic dataset with about ~250K data records.

In Empirical Wavelet construction approach, one of the biggest challenges is to distinguish the high-quality seismic records from noisy records with low quality. In our Empirical Wavelet construction, we have designed an algorithm to empirically distinguish the high-quality data from the noisy data based on prior information, including the signal-to-noise ratio, cross-correlation coefficients, waveform misfit and traffic phase information (other phases being too close to the phase of interest) et al. This classification algorithm is able to distinguish noisy (very low quality) seismic record from high-quality seismic record. However, the performance drops down for those records that are in between high quality and very low quality. Seismic data inherits complex features that is very hard to quantify and the large number of conditions that we have to deal with makes it very difficult to characterize the quality of seismic data based on empirical rules.

Supervised machine learning performs a task effectively without using explicit instructions, rather, it relies on the patterns and statistical inference from the data itself. It has been successfully adapted and deployed in many real-world applications that involves very large and complex dataset, such as computer vision [*Rosten and Drummond, 2006; Vedaldi and Fulkerson, 2010; Szegedy et al., 2016*], natural language processing [Manning and Schütze, 1999; Manning et al., 2014], forecasting [*Ahmed et al., 2010*] and

optimizations [*Scholkopf and Smola, 2001; Snoek et al., 2012; Le et al., 2011*]. Supervised machine learning algorithm takes in an input-output pairs and train a model to map the input to the output labels and the constructed model could then be deployed on new inputs to make predictions. Based on the different type of output used in model training, supervised machine learning can be further divided into supervised classification (where the output labels are discrete) and supervised regression (where the output labels are continuous). In our case, the output is either high-quality or low-quality labels, and thus we could construct a binary classification model.

Considering the large volume of seismic data that is available, the number of different phases and different types of seismic waves, including P, SV and SH, manually label every seismic record is almost impossible. To make empirical wavelet construction method applicable to the vast amount of seismic dataset, we have designed a supervised machine learning method to help us classify high quality seismic data from noisy ones.

3.2 Supervised Classification

Supervised classification takes an input (features) and output(labels) pair and try to map the input features into the output labels. In this study, our objective is to use the information that we have documented in Empirical Wavelet construction and map it into the binary "high-quality"/"low-quality" labels, which is a classic binary classification problem. We use manual labels that we have collected in Empirical Wavelet construction as the training dataset to train a binary classification model and apply it to make predictions on unseen dataset.

3.2.1 Data and Feature Exploration

A machine learning model`s ability to accurately classify different labels depends on the correlation between the features and the target labels. A highly correlated feature would significantly improve the performance of the classification model. However, in most real-world cases, such feature does not exist. To boost the performance of the model, many different features with low correlation are usually used.

In Empirical Wavelet construction, we have comprehensively documented the characteristics of each data record, with a total number of 39 features. In the method, we have manually labeled 1.4 million records for the 6 basic phases, including S, Sdiff, SS, SSS, ScS and ScSScS with about 17.85% of all data records being labeled as high-quality data entry. These data are used as the training dataset to train and test our supervised machine learning model.

To understand the importance of each feature in our dataset, we show the correlation of each feature with respect to the target labels. The correlation of the feature with respect to the target label gives us information on the predictive power or importance of the feature. We have identified some of the highest correlated features in Figure 3.1. We use histogram to show the correlation between target labels and individual features, including distance, Signal-to-noise ratio, station latitude, record Gaussian factor, stretching factor, travel time anomaly and back azimuth. The histograms are color coded, with red corresponds to low-quality data and blue-corresponds to high-quality data. The kernel density function is shown on top as curves for each label. The difference we see in the distribution of the feature shows the correlation between the feature and the target label.

Additionally, we have 6 difference phases listed in our dataset. Phase names are categorical features, we use one hot encoding to convert the 'phase name' column into 6 different features (columns), with each being a binary feature. The correlation between each phase and the target label is shown in Figure 3.2.

3.3 Model Training and Evaluation

Since the target labels in our dataset ("high-quality" versus "low-quality" label) is skewed, with only about 17% of all data being labeled as "high-quality", we use precision and recall instead of accuracy to characterize the performance of the trained binary classification model. When the target labels are skewed, precision and recall describe the performance of the model more accurately. The definition of precision defined as:

$$Precision = \frac{true\ positive}{true\ positive + false\ positive} \tag{3.1}$$

where true positive is the number of records that classified as high quality by our empirical rule and labeled as high quality in Empirical Wavelet construction. False positive is the number of records that is classified as high quality by the empirical rule but is labeled as low quality. Similarly:

$$Recall = \frac{true\ positive}{true\ positive + false\ negative} \tag{3.2}$$

where false negative is the number of records that is classified as low quality by the empirical rule but is labeled as high quality in Empirical Wavelet construction. Using the two equations, our empirical classification model scored a precision of 53% and recall of 71%.

To train and evaluate the performance of the model, we randomly select and reserve 20% of our dataset as test dataset and use the rest 80% of data as development

dataset to train and cross-validate the model performance (see Figure 5.3 for the division, where blue box shows the development dataset and purple box shows the test dataset). Since our target labels are skewed, with ~17% of the labels being high quality, we use a stratified 5-fold cross-validation framework to divide the development dataset into training dataset and cross-validation dataset, as shown in Figure 5.3. The 5-fold framework is an effective way to utilize all development dataset to train the model. The stratified 5-fold cross-validation split the development dataset 5 times and each time the development dataset is divided into 5 equal-sized fold. In each split, we assign similar percentage of high-quality labels as the development dataset to make sure each split has similar skewness.

During model training phase, 4 folds (colored as light blue boxes) are used to train the model and the remaining fold (colored as green is used to cross-validate the model performance. A prediction is made on the remaining fold. We have gathered the predicted results from remaining 1-fold across 5 different splits and assembled them into an out of fold prediction dataset (shown as orange boxes), which has the same size of the development dataset. The out of fold prediction set is then used to evaluate the model.

We use Light Gradient Boosting Machine (LightGBM, [*Meng et al.*, 2016; *Ke et al.*, 2017; *Zhang et al., 2017*]) to train the classification model within the framework of Scikit-learn [*Pedregosa et al, 2011*]. LightGBM is a fast, distributed, high performance gradient boosting framework that is based on decision tree algorithms. And it has been proven to be fast and stable with high performance at supervised learning, for both classification and regression tasks.

The objective of the training is to build a classification model that best separate the binary target labels (high quality and low quality). We use the area under the curve value (AUC) of receiver operation characteristic curve (ROC) to represent the performance of the classification model in the training. ROC curve represents the relationship between sensitivity (true positive / (true positive + false negative)) against specificity (true negative / (true negative + false positive)) at various threshold settings and it allows us to evaluate the performance of the model with a single objective value.

Decision tree-based model are very easy to overfit the data, so necessary steps need to be implemented to avoid overfitting. We have limited the maximum depth of the tree and also limited the number of leaves in one tree. Also, we have setup an early stopping rule where the learning process stops if the AUC ROC score does not improve in 100 iterations. To further boost the performance of the model, we enabled bagging and boosting operation during the training. Bagging is useful as it decreases the model`s variance, where samples are drawn from the training set with replacement [*Breiman*, 1996]. Boosting is another technique that is used to decrease the bias of the model, it uses a set of week learners (decision trees) to create a strong learner by adjusting the weights of data that is misclassified in each subsequent iteration [*Freund and Schapire*, 1997].

We train the model on all development dataset and used the final model to make predictions on the test dataset. By comparing the predicted labels from our classification model and the actual labels from the dataset, we can than evaluate the performance of the model.

Table 3.1 shows the performance of the empirical rule algorithm that is implemented in Empirical Wavelet construction and the performance of LightGBM

model. With the supervised classification using LightGBM, we have improved the precision from ~53% to 87% and recall from 71% to 83%, which is a significant improvement.

3.4 Discussion and Implications

In Empirical Wavelet construction, we have developed an automated data processing algorithm that automatically characterize and document the seismic data measurements and thus applicable to large volume of global seismic dataset. However, to utilize these documented datasets, we need to have an effective approach to classify high-quality seismic records from low quality seismic records. Due to the complex nature of seismic data, accurately classify seismic data remains as a very challenge task and is mostly done through manual labeling [*Astiz et al.*, 2014]. In Empirical Wavelet construction, we have manually labeled more than 1.4 million seismic records together with an empirical rule-based classification model. However, the performance of the model is not good enough to be implemented directed and we only used it as a guidance to help us label the data. The large volume of labeled seismic data together with the rich information that we have documented allows us to train a machine learning model to more accurately implement the classification task.

In this study we have used seismic information documented from Empirical Wavelet construction to train a supervised classification machine learning model to distinguish high quality seismic records from low quality seismic records. We show the correlation between individual feature and the target labels, which presents confidence to build a high-performance classification model. With the dataset, we have used a stratified 5-fold cross-validation scheme to train the model and tested on the reserved test dataset.

Our final model shows significant improvement in both precision and recall for the binary classification task with precision close to 87% and recall close to 83%.

Considering the large volume of seismic records available in global seismic dataset, our final model would still result in 10% to 15% misclassification, where the sheer number of records can be pretty significant. The misclassifications usually introduce large measurement errors and therefore add noise to the predicted "high-quality" dataset. While in practice, together with the binary classification labels, we have also preserved the probability prediction value associated with each label (which has a value between 0.0 and 1.0 where if it is greater than 0.5, it is labeled as high-quality data, otherwise low-quaintly data). The probability prediction associated with each record provides constrain on the confidence of the prediction decision. It could be used as a weight to punish those that have low confidence (with value close to 0.5). Together with the comprehensive weight that we have introduced in Empirical Wavelet construction, we can utilize the constructed dataset with great confidence.

The supervised classification model provides an approach to extract the high confidence high quality seismic records from large volumes of global seismic dataset. Combined with Empirical Wavelet construction, we have the potential to characterize, document and study seismic data through fully automated manners, which opens new ways to study the patterns and anomalies in seismic data.

3.5 References

Ahmed, N. K., Atiya, A. F., Gayar, N. E., & El-Shishiny, H. (2010). An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, *29*(5-6), 594-621.

Astiz, L. et al. (2014), The Array Network Facility Seismic Bulletin: Products and an Unbiased View of United States Seismicity, *Seismological Research Letters*, *85*(3), 576–593, doi:10.1785/0220130141.

Bontempi, G., Taieb, S. B., & Le Borgne, Y. A. (2012, July). Machine learning strategies for time series forecasting. In *European business intelligence summer school* (pp. 62-77). Springer, Berlin, Heidelberg.

Breiman, L. (1996), Bagging predictors, *Mach Learning*, *24*(2), 123–140, doi:10.1007/BF00058655.

Freund, Y., and R. E. Schapire (1997), A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting, *Journal of Computer and System Sciences*, *55*(1), 119–139, doi:10.1006/jcss.1997.1504.

Ke, G., Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu (2017), LightGBM: A Highly Efficient Gradient Boosting Decision Tree, *papers.nips.cc*, 3146–3154.

Lai, H., E. J. Garnero, S. P. Grand, R. W. Porritt, and T. W. Becker (2019), Global Travel Time Data Set From Adaptive Empirical Wavelet Construction, *Geochem. Geophys. Geosyst.*, *20*(5), 2175–2198, doi:10.1029/2018GC007905.

Le, Q. V., Ngiam, J., Coates, A., Lahiri, A., Prochnow, B., & Ng, A. Y. (2011, June). On optimization methods for deep learning. In *Proceedings of the 28th International Conference on International Conference on Machine Learning* (pp. 265-272). Omnipress.

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. MIT press.

Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., & McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations* (pp. 55-60).

Meng, Q., G. Ke, T. Wang, W. Chen, Q. Ye, Z.-M. Ma, and T.-Y. Liu (2016), A Communication-Efficient Parallel Algorithm for Decision Tree, *papers.nips.cc*, 1279–1287.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *Journal of machine learning research*, *12*(Oct), 2825-2830.

Rosten, E., & Drummond, T. (2006, May). Machine learning for high-speed corner detection. In *European conference on computer vision* (pp. 430-443). Springer, Berlin, Heidelberg.

Russell, S. J., and Norvig, P. (2016). Artificial intelligence: a modern approach. Malaysia; *Pearson Education Limited.*

Scholkopf, B., & Smola, A. J. (2001). *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press.

Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms. In *Advances in neural information processing systems* (pp. 2951-2959).

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2818-2826).

Vedaldi, A., & Fulkerson, B. (2010, October). VLFeat: An open and portable library of computer vision algorithms. In *Proceedings of the 18th ACM international conference on Multimedia* (pp. 1469-1472). ACM.

Zhang, H., Si, S., & Hsieh, C. J. (2017). GPU-acceleration for Large-scale Tree Boosting. *arXiv preprint arXiv:1706.08359*.

Table 3.1. Performance comparison between empirical classification algorithm and LightGBM on precision and recall.

Model Name	Precision	Recall
Empirical Algorithm	52.93%	71.68%
LightGBM	87.39%	83.52%

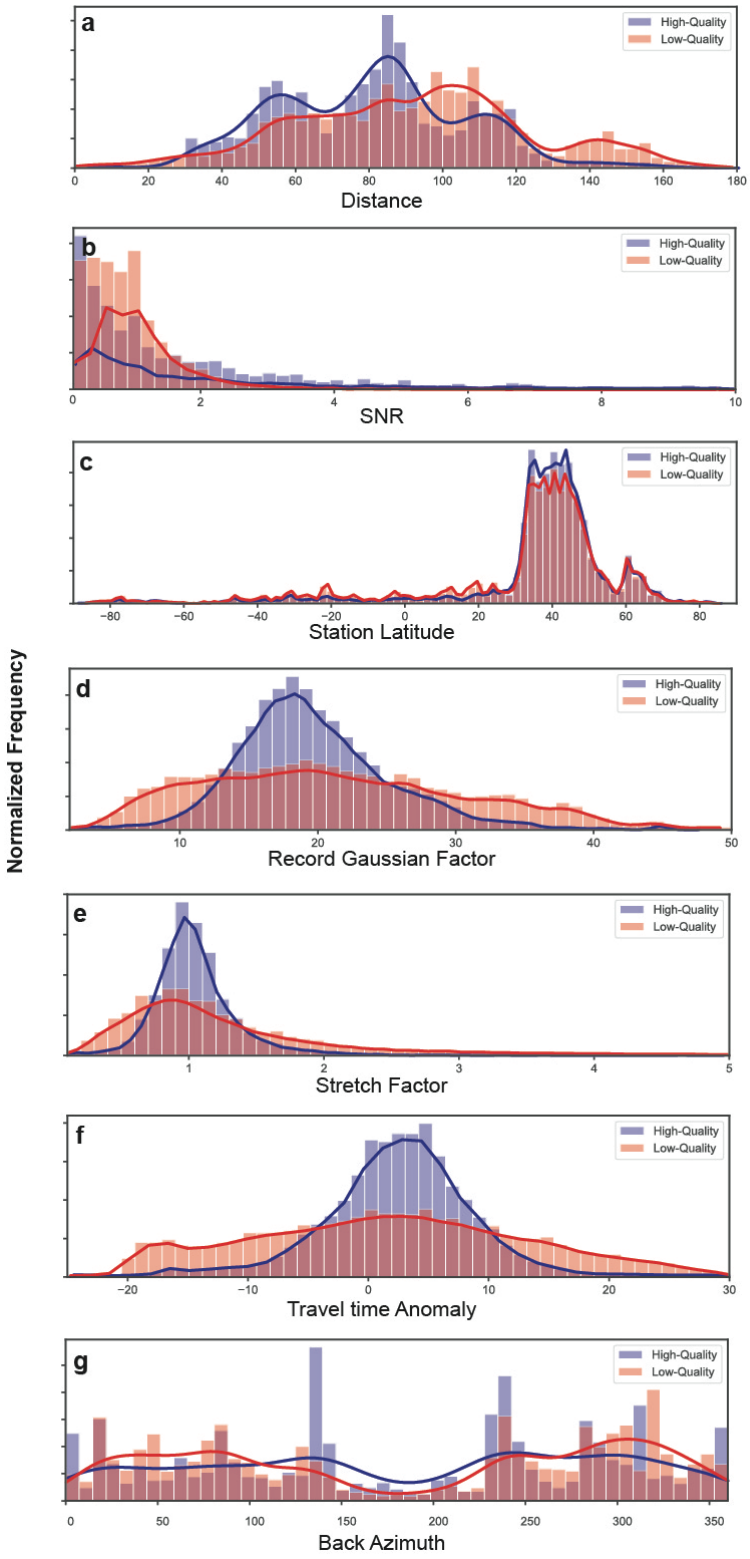
a
High-Quality
Low-Quality
0
20
40
60
80
100
120
140
160
180
Distance
b
High-Quality
Low-Quality
0
2
4
6
8
10
SNR
c
High-Quality
Low-Quality
−80
−60
−40
−20
0
20
40
60
80
Station Latitude
Normalized Frequency
d
High-Quality
Low-Quality
10
20
30
40
50
Record Gaussian Factor
e
High-Quality
Low-Quality
1
2
3
4
5
Stretch Factor
f
High-Quality
Low-Quality
−20
−10
0
10
20
30
Travel time Anomaly
g
High-Quality
Low-Quality
0
50
100
150
200
250
300
350
Back Azimuth

Figure 3.1. Histogram showing correlation between each feature and target label for a) Distance b) Signal-to-Noise ratio c) Station Latitude d) Record Gaussian factor e) Stretching factor f) Travel time anomaly and g) Back azimuth. High-quality data records are represented by blue bar plot and low-quality records are represented by red bars. And kernel density function for each label are drawn on top of the histogram plot as thick curves.

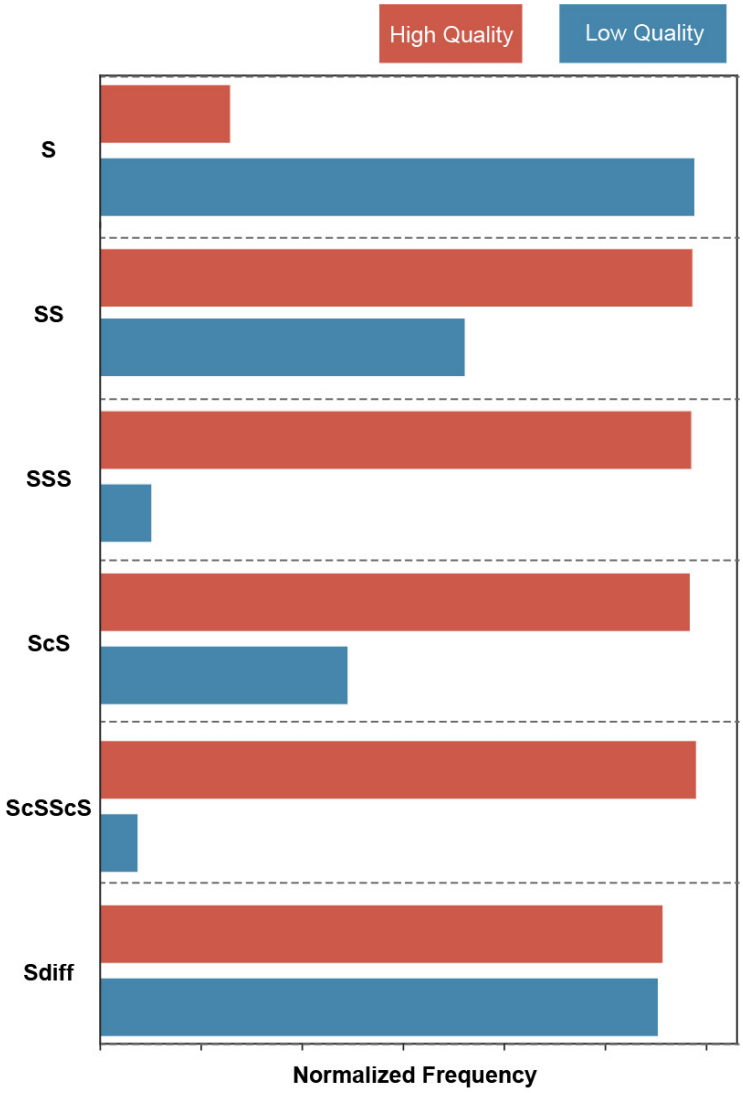

Figure 3.2. Relation between each individual phase (S, SS, SSS, ScS, ScSScS and Sdiff) and the target label (high quality and low quality)

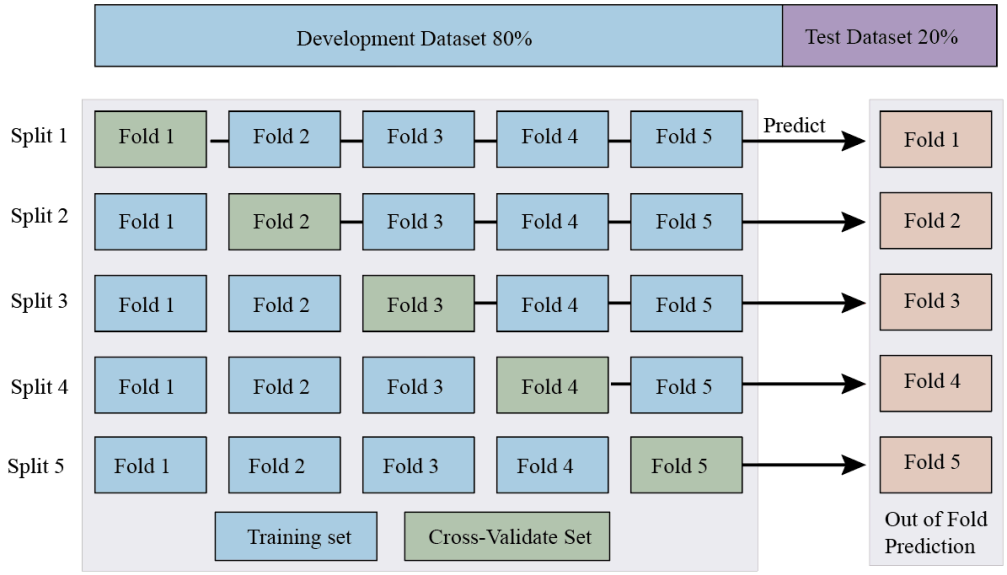

Figure 3.3. Development dataset and test dataset split. 20% of the all dataset is hold out as test dataset and 80% of the dataset is used as development dataset to train and cross-validate the model. The development dataset is further divided into 5 stratified folds, this process is implemented 5 times and each time 4 folds (shows as blue boxes) are used to train the model and 1-fold (shown as green boxes) is used to cross-validate the model. At the same time, the model trained with the 4 folds are used to make a prediction on the remaining fold and all the predicted folds are then assembled into a out of fold prediction, show as orange boxes.

CHAPTER 4

4 TRAVEL TIME AND WAVEFORM MEASUREMENTS OF GLOBAL MULTI-BOUNCE SEISMIC WAVES USING VIRTUAL STATION SEISMOGRAM STACKS

Abstract

We present a method for constructing geographically localized bin stacks of waveforms, here referred to as virtual stations, for the purpose of enhancing signal to noise ratios for travel time and waveform measurements of hard to observe seismic waves. This method is applied to multi-bounce S and ScS phases (S up to S6 and ScS up to ScS5), as well as direct S, ScS, and Sdiff, all on the tangential component of motion. Where visible, major arc S and ScS multi-bounce waves were also measured. Virtual station data are referenced to empirical wavelets constructed from the direct S wave for each event, through an approach to fit the empirical wavelet to the observed wave shapes. The virtual station approach is especially useful for recordings having low signal-to-noise ratios, and provides a way to bolster wave path coverage in the southern hemisphere. Several goodness of fit measurements between the adapted empirical wavelet and virtual station waveforms are documented, as well as signal-to-noise ratios, allowing for objective definition of travel time measurement quality. From a dataset of 360 earthquakes and 8407 seismographic stations, nearly 4 million records were utilized to construct 248,657 virtual station stacked seismograms, which were measured and analyzed in comparison to best-fitting empirical wavelets. After human inspection of virtual station results, 8,871 travel time measurements were retained from 19 different seismic wave types (including minor and major arc paths). Wave path coverage maps show an improved sampling of the southern hemisphere compared to maps without the

higher multiple data. Single seismogram multi-bounce wave measurements were also made, with 3,331 retained travel times from 188,003 records. Comparisons of single seismogram measurements to virtual station stack measurements show a consistent bias: virtual stack onset times are systematically early due to a broadening effect from stacking records with arrival time differences. We report data corrected for this effect, as well as raw data. The travel time dataset, virtual station stacks, and all goodness of fit measurement are publicly available.

4.1 Introduction

Seismic data quantity and type are central to the imaging of Earth's heterogeneous mantle. While surface waves and normal modes are fundamentally important in seismic imaging of the planet, seismic body waves provide the most detailed information of fine scale heterogeneity within the deep interior. For this reason, tomographic imaging routinely employs a large number of body waves of many different seismic phases e.g. [*Becker and Boschi*, 2002; *Grand*, 2002; *Panning and Romanowicz*, 2006; *Houser et al.*, 2008; *Kustowski et al.*, 2008; *Li et al.*, 2008; *Ritsema et al.*, 2011; *Auer et al.*, 2014; *Moulik and Ekström*, 2014; *French and Romanowicz*, 2015; *Koelemeijer et al.*, 2016; *Durand et al.*, 2017]. Comparisons of different global seismic tomography models reveals very similar large-scale structures (~1000 km laterally) in the lower mantle [*Lekic et al.*, 2012; *Cottaar and Lekic*, 2016; *Garnero et al.*, 2016]. However, the intermediate- to small-scale features (at the scale of hundreds of km laterally) vary between models e.g. [*Becker and Boschi*, 2002; *Garnero et al.*, 2016], suggesting uncertainties in the short to intermediate wavelength structure from model to model.

Greater availability of global seismic network data makes possible large data set construction, which may help bolster seismic coverage for some regions. To this end, in a previous study [*Lai et al*., 2019; hereafter Lai19], we used an empirical wavelet construction method to document seismic wave travel time and waveform measurements for the 6 phases S, Sdiff, SS, SSS (S3), ScS and ScSScS (ScS2). An empirical wavelet for each earthquake was iteratively constructed from observed S waves and used as a reference waveform for correlative comparisons with observed waveforms for that event. The method involved adapting the empirical wavelet to best fit every phase of interest, ultimately yielding ~ 250K travel time and waveform measurements for the above 6 phases. However, the wave path sampling of the deep interior from this dataset is significantly better in the northern hemisphere than in the southern hemisphere (roughly 5 times more sampling in the northern hemisphere). Also, many of the data are from common source-receiver geometries (which has not significantly changed over the decades), such as key subduction zones to continents with long-standing seismic networks. As an attempt to mitigate path geometry redundancy, Lai19 prioritized events in unique locations, by pursuing earthquakes with a maximum distance to other earthquakes in the dataset. This still resulted in a sampling bias towards the northern hemisphere.

In an attempt to address the wave path coverage bias, here we investigate multi-bounce seismic waves: S and ScS waves that bounce multiple times off the surface (S and ScS) and core-mantle boundary (ScS). While some of these waves are used in tomographic imaging, e.g., ScS multi-bounce waves up to ScSScSScSScS (ScS4) [*Ritsema et al.*, 2011; *French and Romanowicz*, 2015], such multiple bounce energy is

not currently commonly employed, and higher multiples have not been utilized. Here we pursue any and all seismic body waves apparent on global profiles of transverse component S waves which were not measured in Lai19. These include the major arc SSS (S3m, where major arc wave path departs from the earthquake in the opposite direction from the station along the great circle path), SSSS (S4), SSSS major arc (S4m), SSSSS (S5), SSSSS major arc (S5m), SSSSSS (S6), SSSSSS major arc (S6m), ScSScSScS (ScS3), ScSScSScS major arc (ScS3m), ScSScSScSScS (ScS4), ScSScSScSScS major arc (ScS4m), ScSScSScSScSScS (ScS5), and ScSScSScSScSScS major arc (ScS5m). These phases, especially the major arc phases, can have very long distances and often have paths that traverse the southern hemisphere and provide new and unique wave path sampling locations.

For the phases used in Lai19 (S, SS, S3, ScS, ScS2, Sdiff), an average of roughly 20% of all waves investigated were retained (this varied phase-by-phase), with the high rejection rate being predominantly related to low signal-to-noise ratio (SNR). To utilize more of the available seismic data and successfully pursue the commonly lower amplitude multi-bounce phases, we have adopted stacking-based method to improve the SNR of low amplitude (and sometimes absent) multi-bounce energy. In this paper we present an approach whereby we slowness stack seismic waves in a geographically confined regions, or “bins”. The principle benefit of an array approach is a far improved SNR of waves of interest [*Rost*, 2002; *Frost et al.*, 2013; *Frost and Rost*, 2014], and allows measurements of data that would otherwise be rejected from the single-seismogram approach used in Lai19. We call the geographical bin stacked data in this paper ‘virtual stations’, and measure travel times and waveform information using a

reference empirical wavelet made from stacking S waves on an event-by-event basis (as in Lai19). In addition to employing the virtual station approach on these multi-bounce phases, in this study we also (1) identify wave path corridors for the phases of Lai19 for which there are data but no retained measurements (thus, enhancing SNRs, for unusual and noisy source-receiver geometries), and (2) subsequently investigate all multi-bounce phases for the possibility of single seismogram measurements.

In Section 4.2, we first introduce our data collection and preprocessing procedures. In Section 4.3 we show the workflow of the construction of virtual stations and Section 4.4 presents the virtual station travel time acquisition, measurement trends, single seismogram measurements, and the global coverage of our data. We explore the dependency of our results on some of our assumptions in Section 4.5. Conclusions follow with information on access to the freely available measurements made in this study.

4.2 Global Data Set

4.2.1 Data Collection

In this paper, we investigate earthquakes analyzed in Lai19, so that results of travel time and waveform analyses here can be directly combined with waves of that study (S, SS, S3, ScS, ScS2, Sdiff) from the very same earthquakes. We collected all available data from several data agencies that distribute seismic data freely, including the Incorporated Research Institutions for Seismology (IRIS, http://www.iris.edu), the Observatories & Research Facilities for European Seismology (ORFEUS, http://www.orfeus-eu.org), the Northern California Seismic Network (NECDC, http://www.ncedc.org/ncsn), the F-net Broadband Seismograph Network (F-net, http://www.fnet.bosai.go.jp) and the Canadian National Seismic Network (CNSN,

http://www.earthquakescanada.nrcan.gc.ca/stndon/CNSN-RNSC). A two-step process was employed. First, we collected deep focus events (> 50 km depth) in the time period from January 1994 to October 2017, with moment magnitude greater than 6.0. Smaller moment magnitude earthquakes displayed higher noise levels and inconsistent data quality. Deeper earthquakes tend to have simpler source time functions and less complications from crustal structure (e.g., depth phases and crustal reverberations) interfering with phases of interest. This resulted in 733 earthquakes. However, deep earthquakes are restricted geographically to subduction zones, thus resulting in restricted wave path coverage. Therefore, in a second step we collected select shallow earthquakes (source depth < 50 km) for the same time period, based on a prioritization scheme that (a) favored more recent time periods (to take advantage of more abundantly available data), and (b) ranked events more favorably for having the greatest distance from the deep focus events. The shallow events thus play a fundamentally important role in expanding the global coverage provided solely by the deep earthquakes. This amounted to 113 shallow earthquakes. All events were inspected for possible contamination from other events which occurred nearby in time; if any energy was apparent, the later event was rejected from our catalog. In Lai19, earthquakes were retained if adequate SNR was present such that measurements could be confidently made. This resulted in a final collection of 360 earthquakes (247 deeper than 50 km, 113 shallower than 50 km).

4.2.2 Basic Preprocessing

We collected a 2-hour time window length following the earthquake origin time for all available seismic stations, for all events. For each station, the horizontal components of motion for the broadband data were rotated to the great-circle path to

obtain the radial (R) and transverse (T) components of motion. This paper analyzes transverse components to study SH waves. The instrument response for every station was removed through deconvolution using pole-zero files supplied by the data agency, and data are analyzed on velocity recordings. All data were band pass filtered in the period range between 16 and 100 seconds. This was the same period range as used in Lai19.

An example event that occurred on May 24th, 2013 (latitude: 54.87, longitude: 153.28, source depth: 608.9 km, and moment magnitude: 8.3) is used to show phases we pursue in this study. While this is a large earthquake (the largest in our data set), it is useful because it clearly shows all phases. Figure 4.1 shows almost 2 hours of recordings over the entire distance range. Due to the large number of recordings we have for this event (1801 stations), we linearly stacked records in every 1.0 degree distance window to produce a clearer display of the seismic arrivals that are present. Travel time curves for the PREM model [*Dziewonski and Anderson*, 1981] are also shown. Direct and multi-bounce S waves are seen out to S6 (major arc), and multiple ScS waves are seen out to ScS5. All of the depth phases are also visible, though in this study we do not pursue the depth phases because of the sometimes-complicated wave shapes. This event makes clear the possibility of measuring the higher multi-bounce waves. Figure 4.1a also shows evidence for waves that have traveled in the minor arc direction, and then continue to make an additional round trip (i.e., 360 degrees more). These can be seen between 12 and 30 degrees (minor arc distance) starting at around 5500 seconds. The first of these arrivals is S4+360^{o}. These waves are rare in our data set so we do not include them in our analyses.

To see some of the waves of interest more clearly, Figure 4.1 presents two zoomed in boxes ("Z1" and "Z2") which are shown in Figure 4.2. As in Figure 4.1, these records are distance increment stacked (but with fewer records per stack). In Figure 4.2a, we show distance stacked records from 85-90 degrees where strong S, SS, S3, S4 and ScS2 phases are present. In Figure 4.2b, we show a slightly expanded distance range, but for a much later time window to highlight later arriving multi-bounce energy. Clear major arc multi-bounce waves are seen for S (blue lines: S3m, S4m, S5m, S6m) and minor and major arc multi-bounce waves for ScS (orange lines: ScS4, ScS5, ScS3m, ScS4m, and ScS5m). At distances larger than 95 degrees, there are fewer records per stack, and the seismograms are significantly noisier. This demonstrates the benefit of stacking data to see the higher multiple waves.

The ray paths of the multi-bounce waves pursued in this study are shown in cross-section plots in Figure 4.3, along with the paths of the six phases that were the focus of Lai19. The multi-bounce S*n* and S*n*m waves densely sample the upper part of the mantle while the ScS*n* and ScS*n*m waves sample the entire mantle and provide redundant sampling of the lowermost mantle. Figure 4.3f demonstrates the potential sampling density improvement when incorporating all the multi-bounce data together.

4.3 Virtual Station Seismograms

4.3.1 Development of an Adaptive S-wave Empirical Wavelet

In this study we adopt the method used in Lai19 to construct an average shape of the S-wave, on an event-by-event process. The method is briefly summarized here (and more details and examples can be found in Lai19). First, S-waves are used because they are the first arrival (before core diffraction distances) and have the best SNR compared to

the energy in front of the phase. For each earthquake, we stack all seismograms at distances larger than 30 degrees to avoid complications from upper mantle triplications [*Grand and Helmberger*, 1984; *Song et al.*, 2004; *Tao et al.*, 2017]and smaller than core diffraction to minimize possible attenuation, scattering, or multipathing effects from long paths at the base of the mantle [*Ritsema et al.*, 1998; *Ni et al.*, 2003; 2005]. Waves are initially stacked on the PREM predicted time, and the resulting stack is used to iteratively realign every record to the previous stack (using cross correlation, and the cross-correlation coefficient and SNR are used to weight records in subsequent stack iterations), then re-stacked. When the stack converges to a shape that no longer changes, we define this as the Generalized Empirical Wavelet (GEW). However, the GEW was constructed with some records that are broader and narrower than the GEW. Therefore the shape of the GEW onset is an average shape. To arrive at a more representative shape function, every record is made to fit the GEW by expanding or shrinking it, then the GEW stack is updated once more. Lai19 dubbed this the Stretched Empirical Wavelet (SEW), which had the advantage over the GEW in that it has a more representative onset shape for all records, if the SEW shape is adapted to fit individual records. Thus, for each event, an SEW was constructed for comparison to every wave of interest. The SEW was adapted to fit to each observed wave by either broadening it from convolution with a t* operator [*Futterman*, 1962] (to match records broader than the SEW), or narrowing it (to match records more narrow than the SEW). An onset time is fixed to the adapted SEW through convolution with a Gaussian function, which has the onset time hardwired to the onset of the Gaussian (as defined as the time associated with 0.01 amplitude for the Gaussian peak at 1). The Gaussian function is defined as:

$$G(i) = e^{-\frac{i^2}{2g^2}} \tag{4.1}$$

where i is the number of time points (and thus the length of the function in time points), e is Euler's number, and g is the Gaussian factor (which corresponds to the standard deviation of the function).

This process is automated and provides a stable and objective manner for travel time determination, as well as documentation of wave shape broadness, and other factors (described in Section 4.3.3). Figure 4.4 shows an example of a virtual station stack of a major arc ScS5 (ScS5m, introduced in Section 4.3.3), the best-fitting t*'ed SEW trace, and the best-fitting Gaussian function with the onset time determination. The onset of ScS5m is well captured by this approach.

4.3.2 Virtual Station Grid Parameterization

To set up virtual station stacking, we first defined a grid of equal area cells on Earth's surface. A network of grid points separated by 700 km in latitude and longitude was established, each with a radius of 500 km (thus the entire globe is covered by the grid cells). Several different grid cell separations and radii were investigated, but smaller grid cells resulted in significantly fewer viable virtual stations away from dense seismic networks. Every earthquake and station were assigned to the grid cell within which they were located. Thus, every grid cell has a list of associated events and stations that are within it (if any). Then, for every event-station grid pair, we search the station grid cell for the existence of records for every phase of interest. If the number of records for a phase of interest is greater than a user-defined threshold, then we proceed to construct a virtual station stack for that phase of interest. In this study, if there are 3 or more records

in the grid cell, we proceeded with making the virtual station stack. We experimented with different thresholds. The minimum of 3 was chosen as a compromise between having ample virtual stations globally (larger threshold values limited virtual station global coverage) and not having enough stations populating any grid cell (fewer than 3 commonly yielded noisy stacks). The 360 earthquakes and 8407 unique stations (Figure 4.5a) in our dataset resulted in 289 unique virtual station grid cells (shown as the blue circles in Figure 4.5b). For all the phases of interest in this study (Figure 4.3) and every virtual seismograph possibility for our dataset, the grid center was relocated to be the geographic center of all contributing stations within the initial virtual station grid. Updated grid center locations resulted in roughly ~248K unique virtual station seismograms for our phases of interest. Every virtual station is assigned its own unique identification number, and along with this number we store the seismic phase name, virtual station latitude and longitude, earthquake information, epicentral distance, azimuth and back azimuth. The locations of virtual stations are shown as small black triangles in Figure 4.5b. They differ from each other within a given virtual station cell because each is relocated to be the center of the stations available for the particular phase and earthquake being stacked

We note that for the six phases of Lai19 (S, SS, S3, ScS, ScS2, Sdiff), about 20% of all records analyzed were retained in measurements (roughly 250K measurements were retained from ~1.4 million seismograms investigated). Thus ~80% of the records were rejected, commonly because of low SNR. Therefore we also implement the virtual station approach with these basic 6 phases with the goal of extracting more measurements from the previously rejected records, as they could potentially provide new sampling

geometries. With this goal in mind, we preclude virtual stack construction if we already possess 4 or more measurements for any virtual station geometry for the 6 phases of Lai19. Thus, the virtual station approach for the basic 6 phases is used to fill in unsampled regions and to bolster poorly sampled corridors.

4.3.3 Virtual Station Seismogram Stacking and Measurements

A travel time prediction is computed for the PREM model for every seismogram (and phase of interest) in the retained virtual station cells. For each virtual station cell, phases are aligned on the PREM-predicted time, and weighted according to the distance to the relocated virtual station bin center with a Gaussian factor (a weight of 1 at the bin center that reduces to 0.5 at the edge of the virtual station circle). We utilize focal mechanisms from globalcmt.org [*Dziewonski et al.*, 1981; *Ekström et al.*, 2012] to predict the radiation pattern amplitude of all phases (between -1.0 and 1.0) which is used to predict the polarity of all phases of interest. For any virtual stations possessing records predicted to be in different quadrants of the radiation pattern, records are flipped to the same polarity before virtual station stacking. Finally, all weighted records are stacked. Three examples of virtual station stacking are presented in Figure 4.6. The virtual station construction process is shown for a minor arc ScS4 in Figure 4.6a, for a deep focus Fiji event recorded in northern Japan. The 21 original seismograms are stacked along the PREM predicted ScS4 slowness, resulting in a much-improved SNR (5.0) in comparison to the average SNR of the constituent records (1.2). The SNR method is defined below. A major arc example for ScS5m is presented in Figure 4.6b, for a deep focus South American earthquake recorded at 96 stations in a virtual station cell in the northeastern US. Again, a vast improvement in SNR is apparent (3.6 for the stack

compared to 1.2 for the average of the constituent traces). Lastly, a shallow Alaskan earthquake recorded in a virtual station bin with 80 stations is shown for S5m. Again, the SNR improvement is apparent. The three examples in Figure 4.6 result in a virtual station stack from which onset times are measured with the empirical wavelet method, where by the SEW for each event has been adapted to best bit the virtual station stack (through t* convolution or narrowing pulse width) and a Gaussian has been fit to the stretched SEW which assigns the onset time to the stack (these steps are identical to that in Lai19).

For all virtual station stacks, best-fits of the stretched SEW to the virtual station stack and the Gaussian to the stretched SEW are determined by cross-correlation. We also measure the cross-correlation coefficient between the virtual station stack with the GEW (the unstretched, initial empirical wavelet). These correlation coefficients, along with the measured and predicted arrival times are also stored. We approximate a start and stop time to the arrival of interest by documenting the time associated with 10% amplitude level at the beginning and end of the best-fitting SEW, which we use to estimate wave period.

As noted in Lai19, there are several ways to document the SNR. For virtual stations, we adapted the average amplitude approach, whereby the average amplitude in both a signal window (as defined as one period, approximated as the time between the start and stop time of the phase of interest, previously described) and the noise window (in a window from -100 to -20 sec relative to the PREM predicted time for the phase of interest) are divided. We denote this as SNR_{VS}. We also document the average of the SNRs computed for each record used to create a virtual stack seismogram (notated as $\overline{SNR}$). Additionally, we computed SNR on the virtual stack by dividing by the maximum

peak-to-trough amplitude within one period of the signal (of the phase of interest) and in the noise (in the same 80 second noise window as SNR_{VS}). We notate this as SNR_{peak_trough}. We document if any known seismic waves are predicted to arrive in the time window used to estimate the noise level.

Waveform distortions may occur for some path geometries from a variety of sources, including multi-pathing, where wave energy can take different paths when tangential to or in the vicinity of large-scale heterogeneities and significantly broaden pulses [*Ni*, 2002; *Ni and Helmberger*, 2003a; 2003b; *Ni et al.*, 2005] or scattering from heterogeneity, either fine-scale which can diminish high frequencies, or scales comparable to (or larger than) the dominant seismic wavelength which can result in additional scattered arrivals that can manifest as pre- or post-cursory energy [*Bréger and Romanowicz*, 1998; *Flanagan and Shearer*, 1998; *Rost et al.*, 2008; *Rost and Earle*, 2010; *Rychert and Shearer*, 2010]. We document waveform differences between the observed virtual stack (VS) seismograms and the best-fitting SEW through measurement of the average of their amplitude differences when aligned at their maximum cross-correlation. We define this as misfit:

$$Misfit = \frac{\sum_{i=1}^{n} |A_i^{VS} - A_i^{best-fit\ SEW}|}{n}, \tag{4.2}$$

where A_i^{VS} and $A_i^{best-fit\ SEW}$ are the amplitudes of virtual stack and best-fitting SEW at the i^{th} point, measured across a one period window, and n is the number of points in this window. When measured over one period the phase of interest, we notate this as $Misfit_{SIGNAL}$. We also compute the misfit over one period in the time window immediately preceding and following the signal window, defined as $Misfit_{PRE}$ and $Misfit_{POST}$, respectively.

Lai19 introduced a comprehensive weighting scheme for the basic 6 phases they investigated, as a way to empirically establish comparative data quality between the different phases, for the purpose of future imaging experiments. We use the same approach here, where 5 measured attributes are used as individual weights, which are the SNR_{VS} (w_{SNR}), the cross-correlation coefficient (CCC) between the stretched SEW and the virtual station stack (w_{CCC}), and the misfit measurements of the main phase, pre- and post-cursor time windows ($w_{Mistfit\_signal}$, $w_{Mistfit\_pre}$, and $w_{Mistfit\_post}$, respectively), then multiplied to define a comprehensive weight $w_{comprehensive}$:

$$w_{comprehensive} = w_{SNR} \times w_{CCC} \times w_{Misfit\_signal} \times w_{Misfit\_pre} \times w_{Misfit\_post} \tag{4.3}$$

The weighting factors on the right side of Equation (4.3) are as defined in Figure 4.13 of Lai19, and kept the same so the dataset presented in this paper can easily incorporated with that of Lai19 using the same comprehensive weight values. We briefly reiterate those functions here: w_{SNR} is 1 for SNR≥5, and linearly decreases to 0.5 at SNR=2, and is fixed at 0.5 for SNR<2; w_{CCC} is 1 for CCC≥0.98, and linearly decreases to 0.5 at CCC=0.92, and is fixed at 0.5 for CCC<0.92; $w_{Mistfit\_signal}$ is 1 for $Misfit_{SIGNAL}$≤0.05, and linearly decreases to 0.5 at $Misfit_{SIGNAL}$ =0.30, and is fixed at 0.5 for $Misfit_{SIGNAL}$ >0.30; $w_{Mistfit\_pre}$ is 1 for $Misfit_{PRE}$≤0.10, and linearly decreases to 0.5 at $Misfit_{PRE}$ =0.20, and is fixed at 0.5 for $Misfit_{PRE}$ >0.20; and $w_{Mistfit\_post}$ is 1 for $Misfit_{POST}$≤0.50, and linearly decreases to 0.2 at $Misfit_{POST}$ =1.0, and is fixed at 0.2 for $Misfit_{POST}$ >1.0. While empirically developed, this weighting factor presents a simple approach of comparing and ranking measurements made in this study.

4.3.4 Quality Control

After implementing virtual station stacking for all multi-bounce phases and the phases of Lai19 for the 360 events, 248,657 virtual stations stacks were constructred (~181,451 for new multi-bounce phases introduced here, 67,206 for the 6 basics phases of Lai19). The number of seismograms used in this process was 3,961,572. The multi-bounce phases have significantly longer travel paths in the mantle, and thus are lower amplitude and more attenuated than direct waves. We thus routinely found lower SNR for multi-bounce waves than for direct waves. Here, we followed the approach of Lai19 and constructed PDF files displaying all virtual station seismograms with the best-fitting SEW (and the Gaussian that best-fits the SEW) plotted on top of the phase of interest. This was done on an earthquake-by-earthquake, and phase-by-phase basis. Sixteen virtual station seismograms were plotted per PDF page, which allows simple and effective user interactive reviewing of the virtual station stacks, SEWs, and onset time estimations. We empirically determined if a virtual station record should be retained or discarded, based upon SNR_{VS}, the cross-correlation coefficients between the virtual station stack and the best-fitting SEW and Gaussian, and two misfit measurements ($Misfit_{SIGNAL}$, $Misfit_{PRE}$). Records slated to be retained are shown in the PDF with a red "X" in an interactive box to the right of the waveform; rejected records are shown with the box left unchecked. Our algorithm does not select traces if predictions for interfering phases arrive less then 25 sec from the PREM-predicted time. Since records are slowness stacked on the phase of interest, many interfering phases do not stack coherently, so the user can update the selection to be retained. An additional box to the left of the waveform can be used for alphanumeric input, if needed. The retained or rejected choices made by the algorithm can be updated by the reviewer of the files, then saved, and the choices subsequently

extracted from the PDF files by scripts. Figure 4.7 shows an example of part of a page from a PDF catalog for S4m.

Examples of empirical wavelets fitting multi-bounce waves are shown in Figure 4.8 for all phases studied here that were not measured in Lai19. To objectively present virtual station stack quality, 10 traces were randomly chosen from the population of each wave type, then 5 traces were chosen of the 10 for display in the figure. While the SNR is variable across the virtual station stacks, as is the number of contributing records to each stack, the resulting phase of interest is clear and the SEW matching is robust.

4.4 Results: Travel times, Trends, and Wave Path Coverage

4.4.1 Virtual Station Travel Time Dataset

Here we present the travel time dataset that we have constructed with virtual station seismograms. Measurements were made for 19 distinct seismic phases: S, Sdiff, SS, SSm, S3, S3m, S4, S4m, S5, S5m, S6m, ScS, ScS2, ScS3, ScS3m, ScS4, ScS4m, ScS5, and ScS5m. Roughly 250K virtual station seismograms were constructed, then algorithmically retained or rejected, then subsequently reviewed by the authors to validate or update the retain/reject choices. After this process, 8,871 virtual station seismograms were retained (~ 3.6% retention rate). This number is significantly smaller than the average acceptance rate for the 6 basic phases of Lai19, which is around 20%. However, the Lai19 retention rate was dominated by direct S waves (slightly less that ½ of their dataset of ~252K retained measurements), in which 56.8% of viewed S waves were retained. The number of retained virtual stack seismograms are listed in Table 4.1. As mentioned earlier, longer wave paths (e.g., higher multiple major arc phases) are lower amplitude from geometric spreading and attenuation, thus typically have lower success rates. In the case of SSm, it has fewer possibilities for analysis due to its distance range limit, which is roughly a 20 degree window from 160 to 180 degrees (minor arc distance, thus 180 to 200 degrees major arc distance); larger distances result in diffraction (SdiffSdiff), which was not pursued here, primarily because the ambiguity in the location of diffraction. In general, larger events have larger acceptance rates (Figure 4.9), though variability is apparent (and expected) due to differences in radiation patterns to favorable path geometries with abundant stations.

The number of records used in each virtual station stack is variable, from the minimum number of acceptable records (3 records) up to a maximum of 331stations. Figure 4.10 presents a histogram of the frequency distribution of the number of records used in making virtual station stacks. For locations possessing dense networks (e.g., EarthScope's USArray, see http://earthquake.org), the number of records used is large. For example, 626 of the virtual station stacks have over 100 contributing records. Roughly 43% of the virtual station stacks have between 3 and 10 contributing records, and ~62% of the virtual station stacks have between 3 and 20 contributing records. The SNR of the virtual station stacks is larger than that of the contributing records for the vast majority of our data (Figure 4.11). Data points in Figure 4.11 were computed for virtual stacks made with 3 or more contributing records. While significant scatter is present in the plot, the SNR of virtual stacks are to first order 3 times larger than that of the average SNR of all stations contributing to the virtual station stack.

4.4.2 Single-Seismogram Multi-Bounce Wave Measurements and Virtual Stack Corrections

After visual inspection of summary data record sections for all events, single seismogram multi-bounce waves were apparent for several earthquakes. As a follow-up step, we processed raw single seismogram data in the same fashion as we did with virtual station stacks: a best-fitting SEW was fit to all visible multi-bounce single seismogram phases (as in Section 4.3.1), PDF files were made displaying the best-fitting SEW and Gaussian, then the human review process was conducted for retaining or rejecting data (as in Section 4.3.4). Table 4.2 presents the retained single seismogram measurement counts.

The 3,331 retained single seismogram picks were from records used in virtual station stacks, and thus allow us to compare single-seismogram measurements to the stacks they contributed to. For virtual station stacks having 5 or more single seismogram measurements, the onset travel times are averaged and compared to the onset time measured for the virtual station stack in Figure 4.12a (for all measured phases). A least squares best-fit line is also shown (with an R^2 value of 0.86), and indicates that virtual station stack onset time determinations are several seconds earlier than the average of the onset times of single seismograms measured for that same bin, but well correlated. This is expected – the onset time of the virtual station stack is influenced by the earliest arriving phases in the stack, over the average onset time of records contributing to the stack. In Figure 4.12b we compare the best-fit Gaussian of virtual station stacks to the average of the best-fit Gaussians of the single seismograms that contributed to each virtual station Gaussian; each was referenced to the Gaussian of the unstretched SEW for each earthquake (this latter step was taken to remove the effect of event size so data are

comparable from different events). This comparison highlights a trend: the Gaussians of virtual station stacks are broader if the contributing stations have relatively broadened best-fit Gaussians. This is also expected, for a given distribution of travel time perturbations in a virtual stack bin, broader contributing records will result in a broader stack. The trend lines in Figure 4.12a and 4.12b were weighted using a combination of the SNR of the contributing stations and the SNR of the virtual station stack. We did not find any dependency on specific seismic phases, thus the trendline in Figure 4.12b can be used to estimate, from virtual station Gaussians, an estimation of the average Gaussian factor for the contributing stations. Therefore, arrival time can be estimated from the scaled Gaussian (as in Section 4.3.1), which will more appropriately depict the average time of records contributing to the virtual station stack. We thus develop the correction to achieve this outcome, and indeed the corrected virtual station stack times agree well with the average time anomalies of the contributing stations (Figure 4.12c). We apply this correction to all virtual station times.

For all retained measurements, an ASCII file is made available with measurements and measured attributes made here, including the measured and corrected travel times relative to the PREM model. This file is freely shared via Zenodo (doi:10.5281/zenodo.3247094). Table 4.3 presents the tabulated information descriptions for the virtual station stack measurements file, and Table 4.4 presents the information descriptions for the single seismogram measurements file.

4.4.3 Wave Path Sampling Coverage

Virtual station construction results in a stack with an improved SNR over single seismograms, making possible measurements from multi-bounce phases that are otherwise difficult to investigate, resulting in new path geometries, especially with the major arc data. In Figure 4.13 we present wave path sampling as a function of latitude, by counting rays in 5x5 degree cells. The coverage in the upper mantle (0-660 km, top row of panels) is compared to the deepest 300 km of the mantle (2591-2891 km, bottom row of panels). We first show the latitudinal sampling of the six phases of Lai19 (first column). There is a clear bias of greater sampling in the northern hemisphere, which is dominantly due to the large number of S waves in that dataset, and the predominance of stations in the northern hemisphere. The coverage of all virtual station stacks measured here are shown in the second column. The latitudinal sampling is significantly improved in the southern hemisphere, particularly in the upper mantle. The third column shows the coverage from virtual station stacks of multi-bounce S waves (Sn, n=2-6, minor and major arc). These improve sampling in both hemispheres. The multibounce ScS wave coverage (ScSn, n=2-5, minor and major arc), are presented in the fourth column, and only slightly improve southern hemisphere sampling. This is primarily due to the very low number of major arc ScSn measurements, and the fact that the higher multiple ScSn commonly sample the same corridors of ScS (i.e., a single reflection, n=1). Latitudinal coverage for all minor arc virtual station stacks are shown in the 5th column and is similar (but slightly improved) to the ScSn coverage. However, if only major arc virtual station paths are considered (6th and final column) we see the best relative southern hemisphere sampling. This highlights the potential benefit of major arc paths in whole mantle imaging.

Lateral ray path coverage maps for all virtual station paths for the upper 660 km and deepest 300 km of the mantle are presented in Figure 4.14a. Both depth shells have reasonably good coverage (with the upper mantle being better sampled). Figure 4.14b shows a sampling coverage density map, in 5×5 degree cells for the same depth shells. While it is clear that the upper mantle has slightly greater sampling, as suggested in Figure 4.13, the southern hemisphere is fairly well sampled. To illustrate the sampling density improvement when using higher multiple bounce waves (especially major arc data) investigated here, compared to phases used in Lai19 (S, SS, S3, Sdiff, ScS, ScS2), we present a representation of sampling density improvement in Figure 4.14c. Here, improvement is defined as percentage increase of sampling in the cell by the addition of the virtual station data (as the number of virtual station paths divided by the number of paths in the Lai19 dataset, times 100). The warmer colors mark the most improved regions, with oarange and red marking more than a 100% improvement or more. The southern hemisphere coverage improves significantly from the virtual station stack data (especially the major arc data).

4.5 Discussion

In this paper we present a virtual station stacking method to exploit the use of multi-bounce data that is otherwise typically too low in amplitude to be measured on single seismogram. The virtual station stacking increases the SNR, and combined with the empirical wavelet approach, provides an objective method to measure travel time and waveform information. Some systematic broadening was apparent in the virtual station stacks and corrected for. This method was deployed on all measurable phases in our data set. Single seismograms were also measured where possible, and all measurements were

visually inspected. Figure 4.12 highlights the fact that virtual station stacks are broadened compared to the constituent seismograms (where measurable). A correction for this effect allows travel time estimates of the virtual stacks. An empirical comprehensive weight (Equation 4.3) permits a relative weight for travel times presented here.

However, even with an improved SNR over constituent contributing seismograms, virtual stations average typically noisy seismograms within a geographical bin, and uncertainties exist. For example, we chose to stack stations within a geographic bin having a 5 degree radius from the grid center, with a Gaussian weighting (G) relative to the grid center:

$$G(R) = e^{-\frac{R^2}{2g_R^2}} \tag{4.4}$$

Here, R is radius in degrees. We used a Gaussian factor of g_R=4, which gives $G(5)=0.5$, thus the weighting tapers from 1 at the grid center ($R=0$) to 0.5 at $R=5$. The large stacking radius was chosen to include more stations in order to permit more virtual station stacks, as well as an attempt to improve the SNR. We present the effect of different g_R factors on the radius weighting in Figure 4.15a. Four different g_R values are shown. The effect of different g_R is shown for a multi-bounce ScS wave example (virtual station stack of ScS3m, made from 24 contributing records) in Figure 14.5b. The virtual station stacks for the same four g_R values are presented and show the coherent stacking of ScS3m for $g_R \geq 2$. Larger g_R results in improved SNR, and while g_R=6 has slightly better SNR than g_R=4, we choose the latter to minimize the blurring effect of the greater weight for the more distant stations contributing to the stack. This is apparent for g_R=6 from the slightly larger standard deviation. We note the onset time of the different virtual station stacks do not change for the different g_R.

While the onset assignment to data using the empirical wavelet method is objective, and uncertainties in onset time estimations are likely within ± 1 sec for the wavelet at hand, the broadened virtual station stacks relative to individual stations present additional uncertainties (as detailed in Figure 4.12). While we correct for the wave broadening to shift the onset time estimations, there is some scatter in the comparison of the averaged onset times of constituent records compared to the corrected virtual station stack onset times (Figure 4.12c). Several seconds of scatter is apparent, and sometimes larger. We have omitted virtual station stacks if their onset times differ from the average of onset times of contributing station averages by more than 8 seconds (the ½ period of the upper corner of the bandpass filter used on the data, and the average period of most of the data) in the line fitting in Figure 4.12; this should preclude phase misidentification. We emphasize the importance of the comprehensive weight we give to all measurements, which depends upon factors that characterize SNR and goodness of fit of the SEW to the phases of interest.

We further note that our coverage discussions (e.g., Figures 4.13 and 4.14) were based upon infinite frequency ray paths. The average period of minor and major arc multi-bounce wave virtual station stacks measured here are 17.6 and 22.5 sec, respectively. These can be considered small to intermediate period. While ray computations for coverage may be a reasonable approximation for this period, the sensitivity of these waves, especially those with long paths (e.g., all major arc phases), spans a volume which should be taken into consideration for imaging purposes. Other corrections were not applied here and should be similarly taken into consideration, like

crustal corrections (e.g., [*Artemieva and Mooney*, 2001]) and ellipticity corrections ([*Kennett and Gudmundsson*, 1996]).

4.6 Conclusions

We presented a virtual station stacking algorithm that computes a seismogram stack for geographical bins to improve signal to noise ratios of typically low amplitude seismic waves. The main focus of this study was multi-bounce waves not measured in Lai19, namely S4, S5 (and major arc counterparts S2m, S3m, S4m, S5m, and S6m), ScS3, ScS4, and ScS5 (and major arc counterparts ScS3m, ScS4m, ScS5m). We also constructed virtual station stacks for phases of Lai19 (S, SS, S3, ScS, ScS2, and Sdiff) for poorly or un-sampled corridors. For a dataset of 360 global earthquakes and 8407 seismographic stations, we defined 289 virtual station grid cells, which were used to compute 248,657 virtual station stacks of the 19 different seismic phases. After visual inspection, 8,871 of these were retained. Onset travel time and waveform information was retained. We also searched for single seismogram multi-bounce waves which were present for some stronger earthquakes, and identified and measured 3,331 records. These measurements were compared to the virtual station stacks, and used to develop a wave shape width correction to the virtual station stacks, which are broadened by summing data with time shifts. This resulted in an onset time correction that was applied to all virtual station stack measurements. The resulting wave path coverage from the multi-bounce data significantly helps to bolster sampling in the southern hemisphere. All measurements and data attributes are made publicly available.

4.7 References

Artemieva, I. M., and W. D. Mooney (2001), Thermal thickness and evolution of Precambrian lithosphere: A global study, *Journal of Geophysical Research: Solid Earth (1978–2012)*, *106*(B8), 16387–16414, doi:10.1029/2000JB900439.

Auer, L., L. Boschi, T. W. Becker, T. N. Meyer, and D. Giardini (2014), Savani: A variable resolution whole-mantle model of anisotropic shear velocity variations based on multiple data sets, *Journal of Geophysical Research: Solid Earth*, *119*(4), 3006–3034, doi:10.1002/2013JB010773.

Becker, T., and L. Boschi (2002), A comparison of tomographic and geodynamic mantle models, *Geochem. Geophys. Geosyst.*, 1–48.

Bréger, L., and B. Romanowicz (1998), Three-Dimensional Structure at the Base of the Mantle Beneath the Central Pacific, *Science*, *282*(5389), 718–720, doi:10.1126/science.282.5389.718.

Cottaar, S., and V. Lekic (2016), Morphology of seismically slow lower-mantle structures, *Geophysical Journal International*, *207*(2), 1122–1136, doi:10.1093/gji/ggw324.

Durand, S., E. Debayle, Y. Ricard, C. Zaroli, and S. Lambotte (2017), Confirmation of a change in the global shear velocity pattern at around 1000 km depth, *Geophysical Journal International*, *211*(3), 1628–1639, doi:10.1093/gji/ggx405.

Dziewonski, A. M., and D. L. Anderson (1981), Preliminary reference Earth model, *Physics of the Earth and Planetary Interiors*, *25*(4), 297–356, doi:10.1016/0031-9201(81)90046-7.

Dziewonski, A. M., T. A. Chou, and J. H. Woodhouse (1981), Determination of earthquake source parameters from waveform data for studies of global and regional seismicity, *Journal of Geophysical Research: Solid Earth (1978–2012)*, *86*(B4), 2825–2852, doi:10.1029/JB086iB04p02825.

Ekström, G., M. Nettles, and A. M. Dziewonski (2012), The global CMT project 2004–2010: Centroid-moment tensors for 13,017 earthquakes, *Physics of the Earth and Planetary Interiors*, *200-201*, 1–9, doi:10.1016/j.pepi.2012.04.002.

Flanagan, M. P., and P. M. Shearer (1998), Global mapping of topography on transition zone velocity discontinuities by stacking SS precursors, *Journal of Geophysical Research: Solid Earth (1978–2012)*, *103*(B2), 2673–2692, doi:10.1029/97JB03212.

French, S. W., and B. Romanowicz (2015), Broad plumes rooted at the base of the Earth's mantle beneath major hotspots, *Nature*, *525*(7567), 95–99, doi:10.1038/nature14876.

Frost, D. A., and S. Rost (2014), The P-wave boundary of the Large-Low Shear Velocity Province beneath the Pacific, *Earth and Planetary Science Letters*, *403*, 380–392, doi:10.1016/j.epsl.2014.06.046.

Frost, D. A., S. Rost, N. D. Selby, and G. W. Stuart (2013), Detection of a tall ridge at the core–mantle boundary from scattered PKP energy, *Geophysical Journal International*, *195*(1), 558–574, doi:10.1093/gji/ggt242.

Futterman, W. I. (1962), Dispersive body waves, *Journal of Geophysical Research: Solid Earth (1978–2012)*, *67*(13), 5279–5291, doi:10.1029/JZ067i013p05279.

Garnero, E. J., A. K. McNamara, and S.-H. Shim (2016), Continent-sized anomalous zones with low seismic velocity at the base of Earth's mantle, *Nature Geosci*, *9*(7), 481–489, doi:10.1038/ngeo2733.

Grand, S. P. (2002), Mantle shear–wave tomography and the fate of subducted slabs, *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *360*(1800), 2475–2491, doi:10.1098/rsta.2002.1077.

Grand, S. P., and D. V. Helmberger (1984), Upper mantle shear structure of North America, *Geophysical Journal International*, *76*(2), 399–438, doi:10.1111/j.1365-246X.1984.tb05053.x.

Houser, C., G. Masters, P. Shearer, and G. Laske (2008), Shear and compressional velocity models of the mantle from cluster analysis of long-period waveforms, *Geophysical Journal International*, *174*(1), 195–212, doi:10.1111/j.1365-246X.2008.03763.x.

Kennett, B. L. N., and O. Gudmundsson (1996), Ellipticity corrections for seismic phases, *Geophysical Journal International*, *127*(1), 40–48, doi:10.1111/j.1365-246X.1996.tb01533.x.

Koelemeijer, P., J. Ritsema, A. Deuss, and H. J. Van Heijst (2016), SP12RTS: a degree-12 model of shear- and compressional-wave velocity for Earth's mantle, *Geophysical Journal International*, *204*(2), 1024–1039, doi:10.1093/gji/ggv481.

Kustowski, B., G. Ekström, and A. M. Dziewonski (2008), Anisotropic shear-wave velocity structure of the Earth's mantle: A global model, *Journal of Geophysical Research: Solid Earth (1978–2012)*, *113*(B6), B06306, doi:10.1029/2007JB005169.

Lekic, V., S. Cottaar, A. Dziewonski, and B. Romanowicz (2012), Cluster analysis of global lower mantle tomography: A new class of structure and implications for chemical heterogeneity, *Earth and Planetary Science Letters*, *357-358*, 68–77, doi:10.1016/j.epsl.2012.09.014.

Lai, H., E. J. Garnero, S. P. Grand, R. W. Porritt, and T. W. Becker (2019), Global Travel Time Data Set From Adaptive Empirical Wavelet Construction, *Geochem. Geophys. Geosyst.*, *20*(5), 2175–2198, doi:10.1029/2018GC007905.

Li, C., R. D. van der Hilst, E. R. Engdahl, and S. Burdick (2008), A new global model for P wave speed variations in Earth's mantle, *Geochem. Geophys. Geosyst.*, *9*(5), n/a–n/a, doi:10.1029/2007GC001806.

Moulik, P., and G. Ekström (2014), An anisotropic shear velocity model of the Earth's mantle using normal modes, body waves, surface waves and long-period waveforms, *Geophysical Journal International*, *199*(3), 1713–1738, doi:10.1093/gji/ggu356.

Ni, S. (2002), Sharp Sides to the African Superplume, *Science*, *296*(5574), 1850–1852, doi:10.1126/science.1070698.

Ni, S., and D. V. Helmberger (2003a), Further constraints on the African superplume structure, *Physics of the Earth and Planetary Interiors*, *140*(1-3), 243–251, doi:10.1016/j.pepi.2003.07.011.

Ni, S., and D. V. Helmberger (2003b), Ridge-like lower mantle structure beneath South Africa, *J. Geophys. Res.*, *108*(B2), 3433–14, doi:10.1029/2001JB001545.

Ni, S., D. V. Helmberger, and J. Tromp (2005), Three-dimensional structure of the African superplume from waveform modelling, *Geophysical Journal International*, *161*(2), 283–294, doi:10.1111/j.1365-246X.2005.02508.x.

Ni, S., V. F. Cormier, and D. V. Helmberger (2003), A Comparison of Synthetic Seismograms for 2D Structures: Semianalytical versus Numerical, *Bulletin of the Seismological Society of America*, *93*(6), 2752–2757, doi:10.1785/0120030011.

Panning, M., and B. Romanowicz (2006), A three-dimensional radially anisotropic model of shear velocity in the whole mantle, *Geophysical Journal International*, *167*(1), 361–379, doi:10.1111/j.1365-246X.2006.03100.x.

Ritsema, J., A. Deuss, H. J. Van Heijst, and J. H. Woodhouse (2011), S40RTS: a degree-40 shear-velocity model for the mantle from new Rayleigh wave dispersion, teleseismic traveltime and normal-mode splitting function measurements, *Geophysical Journal International*, *184*(3), 1223–1236, doi:10.1111/j.1365-246X.2010.04884.x.

Ritsema, J., S. Ni, D. V. Helmberger, and H. P. Crotwell (1998), Evidence for strong shear velocity reductions and velocity gradients in the lower mantle beneath Africa, *Geophysical Research Letters*, *25*(23), 4245–4248, doi:10.1029/1998GL900127.

Rost, S. (2002), Array seismology: Methods and applications, *Rev. Geophys.*, *40*(3), 1008, doi:10.1029/2000RG000100.

Rost, S., and P. S. Earle (2010), Identifying regions of strong scattering at the core–mantle boundary from analysis of PKKP precursor energy, *Earth and Planetary Science Letters*, *297*(3-4), 616–626, doi:10.1016/j.epsl.2010.07.014.

Rost, S., E. J. Garnero, and Q. Williams (2008), Seismic array detection of subducted oceanic crust in the lower mantle, *Journal of Geophysical Research: Solid Earth (1978–2012)*, *113*(B6), doi:10.1029/2007JB005263.

Rychert, C. A., and P. M. Shearer (2010), Resolving crustal thickness using SS waveform stacks, *Geophysical Journal International*, *180*(3), 1128–1137, doi:10.1111/j.1365-246X.2009.04497.x.

Song, T.-R. A., D. V. Helmberger, and S. P. Grand (2004), Low-velocity zone atop the 410-km seismic discontinuity in the northwestern United States, *Nature*, *427*(6974), 530–533, doi:10.1038/nature02231.

Tao, K., S. P. Grand, and F. Niu (2017), Full-waveform inversion of triplicated data using a normalized-correlation-coefficient-based misfit function, *Geophysical Journal International*, *210*(3), 1517–1524, doi:10.1093/gji/ggx249.

Table 4.1. Phase-by-phase listing of number of virtual station seismograms retained for measurements, total number of viewed virtual station seismograms, and the ratio of these, i.e., the retention rate.

Phase	Retained	Total viewed	Retention Rate
S	561	2,045	27.43%
Sdiff	438	5,514	7.94%
SS	1,652	19,724	8.37%
SSm	280	2,766	10.12%
S3	1,767	15,011	11.77%
S3m	719	20,434	3.51%
S4	444	6,847	6.48%
S4m	387	25,079	1.54%
S5	142	1,984	7.15%
S5m	83	16,412	0.50%
S6m	41	16,527	0.24%
ScS	464	6,414	7.23%
ScS2	734	18,498	3.96%
ScS3	405	15,980	2.53%
ScS3m	216	11,979	1.80%
ScS4	391	24,737	1.58%
ScS4m	24	12,054	0.19%
ScS5	69	12,371	0.55%
ScS5m	54	14,281	0.37%
TOTAL	**8,871**	**248,657**	**3.57%**

Table 4.2. Phase-by-phase listing of number of single seismogram measurements retained, viewed, and the retention rate of each phase.

Phase	Retained	Total viewed	Retention Rate
S3m	879	36,298	2.42%
S4m	226	35,165	0.64%
S5m	22	11,688	0.19%
S6m	4	2,252	0.18%
ScS3	1,678	39,364	4.26%
ScS4	457	36,787	1.24%
ScS4m	17	4,721	0.36%
ScS5	48	21,728	0.22%
TOTAL	**3,331**	**188,003**	**1.77%**

Table 4.3. List of virtual station stack measurements and attributes computed in this study that are shared in an archived data file

#	Information	Description
1	VS[a] name	Phase name followed by an integer as a code to represent VS station name
2	Distance	Epicentral distance between event and VS grid center in degrees
3	Stack Number	Number of records that is used to stack for virtual stations
4	VS latitude	VS location latitude in degrees
5	VS longitude	VS location longitude in degrees
6	Event latitude	Earthquake hypocentral location latitude in degrees[b]
7	Event longitude	Earthquake hypocentral location longitude in degrees[b]
8	Event depth	Earthquake hypocentral location depth in km[b]
9	Event magnitude	Earthquake moment magnitude[b]
10	Origin time	Earthquake origin time[b]
11	Azimuth	Azimuth from earthquake to VS grid (in degrees)
12	Back azimuth	Back azimuth measured at VS clockwise back to earthquake (in degrees)
13	Phase name	Either S, Sdiff, SS, SSm, S3, S3m, S4, S4m, S5, S5m, S6m, ScS, ScS2, ScS3, ScS3m, ScS4, ScS4m, ScS5, or ScS5m
14	Predicted time	Travel time prediction of the PREM model
15	Measured time	Travel time anomaly of phase onset relative to PREM (observed minus PREM)
16	Corrected time	Travel time anomaly of phase onset relative to PREM (observed minus PREM) corrected for over-broadened virtual station stack
17	Phase start	The start time, relative to PREM prediction, of the beginning of the time window used to define one pulse width of VS phase of interest on velocity recordings, measured at the 10% amplitude level preceding the wave peak (used to auto-define the Misfit measurement windows)
18	Phase end	The end time, relative to the PREM prediction, of the end of the time window used to define one pulse width of VS phase of interest on velocity recordings, measured at the 10% amplitude level following the wave peak (used to auto-define the Misfit measurement windows)
19	SNR_{VS}	VS SNR measurement from the average amplitude of the signal to the average amplitude of the noise
20	$\overline{SNR}$	Average SNR of all records used to create a VS stack
21	SNR_{peak_trough}	VS SNR measurement from the maximum peak-to-trough value within one period of the signal compared to noise

22	$Misfit_{SIGNAL}$	The average difference between the VS phase and the best-fit SEW over one period (as in Equation 2)
23	$Misfit_{PRE}$	As above, except over one period preceding the phase of interest
24	$Misfit_{POST}$	As above, except over one period following the phase of interest
25	t*	The best-fit t^* value that, when convolved with the SEW, gives the best fit to records that are broader than the SEW
26	Stretch factor	A measure of the amount the SEW has to be narrowed to fit records that are narrower than the SEW
27	$CCC_{[rec,SEW]}$	Cross-correlation coefficient between observed record and the best-fitting SEW adapted to the record
28	$CCC_{[rec,GEW]}$	Cross-correlation coefficient between observed record and the GEW, which measures the record's fit to the average S wave phase shape for the event
29	$g_{\_best\text{-}fit\_SEW}$	Gaussian factor of the best-fitting Gaussian function (g, Equation 1) to a record's best-fitting SEW
30	$g_{\_event\_SEW}$	Gaussian factor of the best-fitting Gaussian function (g, Equation 1) to the GEW for the event
31	$Misfit_g$	The misfit measured between $g_{\_best\text{-}fit\_SEW}$ and $g_{\_event\_SEW}$ (computed as in Equation 2) which provides a different measure of record broadening
32	$w_{\_comprehensive}$	An empirical comprehensive weight value for each data (see Equation 3)
33	Noise window traffic flag	Records that have "traffic" (interfering seismic waves) predicted to arrive in the noise window (of the SNR measurement) are flagged as 1, otherwise 0
34	Period	Estimated period of the phase, from the start and end of the pulse measured at 0.1 amplitude (when peak is set to 1) measured on displacement recordings
35	Polarity	The predicted amplitude between [−1,1] using the SH radiation pattern for the phase of interest

Note. The number in the first column of the table below corresponds to the column number in archived ASCII file.

[a]Virtual Station.

[b]As provided by ISC (International Seismological Center)

Table 4.4. List of single seismogram measurements and attributes computed in this study that are shared in an archived data file

#	Information	Description
1	Station name	The 3-5 character station name code
2	Network name	The two-digit code for the seismographic network
3	Distance	Epicentral distance between earthquake and station in degrees
4	Station latitude	Station location latitude in degrees[a]
5	Station longitude	Station location longitude in degrees[a]
6	Event latitude	Earthquake hypocentral location latitude in degrees[b]
7	Event longitude	Earthquake hypocentral location longitude in degrees[b]
8	Event depth	Earthquake hypocentral location depth in km[b]
9	Event magnitude	Earthquake moment magnitude[b]
10	Origin time	Earthquake origin time[b]
11	Azimuth	Azimuth from earthquake to station (in degrees)
12	Back azimuth	Back azimuth measured at station clockwise back to earthquake (in degrees)
13	Phase name	Either S3m, S4m, S5m, S6m, ScS3, ScS4, ScS4m, or ScS5
14	Predicted time	Travel time prediction of the PREM model
15	Measured time	Travel time anomaly of phase onset relative to PREM (observed minus PREM)
17	Phase start	The start time, relative to PREM prediction, of the beginning of the time window used to define one pulse width of phase of interest on velocity recordings, measured at the 10% amplitude level preceding the wave peak (used to auto-define the Misfit measurement windows)
18	Phase end	The end time, relative to the PREM prediction, of the end of the time window used to define one pulse width of phase of interest on velocity recordings, measured at the 10% amplitude level following the wave peak (used to auto-define the Misfit measurement windows)
19	$SNR_{average\_amp}$	The signal-to-noise measurement from the average amplitude of the signal to the average amplitude of the noise
20	$SNR_{peak\text{-}trough}$	The signal-to-noise measurement from the maximum peak-to-trough measurement within one period of the signal compared to noise
21	SNR_{max_peak}	The signal-to-noise measurement from the maximum peak in the signal window compared to the maximum peak in the entire noise window
22	$Misfit_{SIGNAL}$	The average difference between the phase and the best-fit SEW over one period
23	$Misfit_{PRE}$	As above, except over one period preceding the phase

24	Misfit_{POST}	As above, except over one period following the phase
25	t*	The best-fit t^* value that, when convolved with the SEW, gives the best fit to records that are broader than the SEW
26	Stretch factor	A measure of the amount the SEW has to be narrowed to fit records that are narrower than the SEW
27	$CCC_{[rec,SEW]}$	Cross-correlation coefficient between observed record and the best-fitting SEW adapted to the record
28	$CCC_{[rec,GEW]}$	Cross-correlation coefficient between observed record and the GEW, which measures the record's fit to the average S wave phase shape for the event
29	$g_{\_best\text{-}fit\_SEW}$	Gaussian factor of the best-fitting Gaussian function (g, Equation 1) to a record's best-fitting SEW
30	$g_{\_event\_SEW}$	Gaussian factor of the best-fitting Gaussian function (g, Equation 1) to the GEW for the event
31	Misfit_g	The misfit measured between $g_{\_best\text{-}fit\_SEW}$ and $g_{\_event\_SEW}$ (computed as in Equation 2) which provides a different measure of record broadening
32	$w_{\_comprehensive}$	An empirical comprehensive weight value for each data (see Equation 3)
33	Noise window traffic flag	Records that have "traffic" (interfering seismic waves) predicted to arrive in the noise window (of the SNR measurement) are flagged as 1, otherwise 0
34	Period	Estimated period of the phase, from the start and end of the pulse measured at 0.1 amplitude (when peak is set to 1) measured on displacement recordings
35	Polarity	The predicted amplitude between [−1,1] using the SH radiation pattern for the phase of interest

Note. The number in the first column of the table below corresponds to the column number in archived ASCII file.

[a]As provided by the data agencies listed in Section 2.1

[b]As provided by ISC (International Seismological Center)

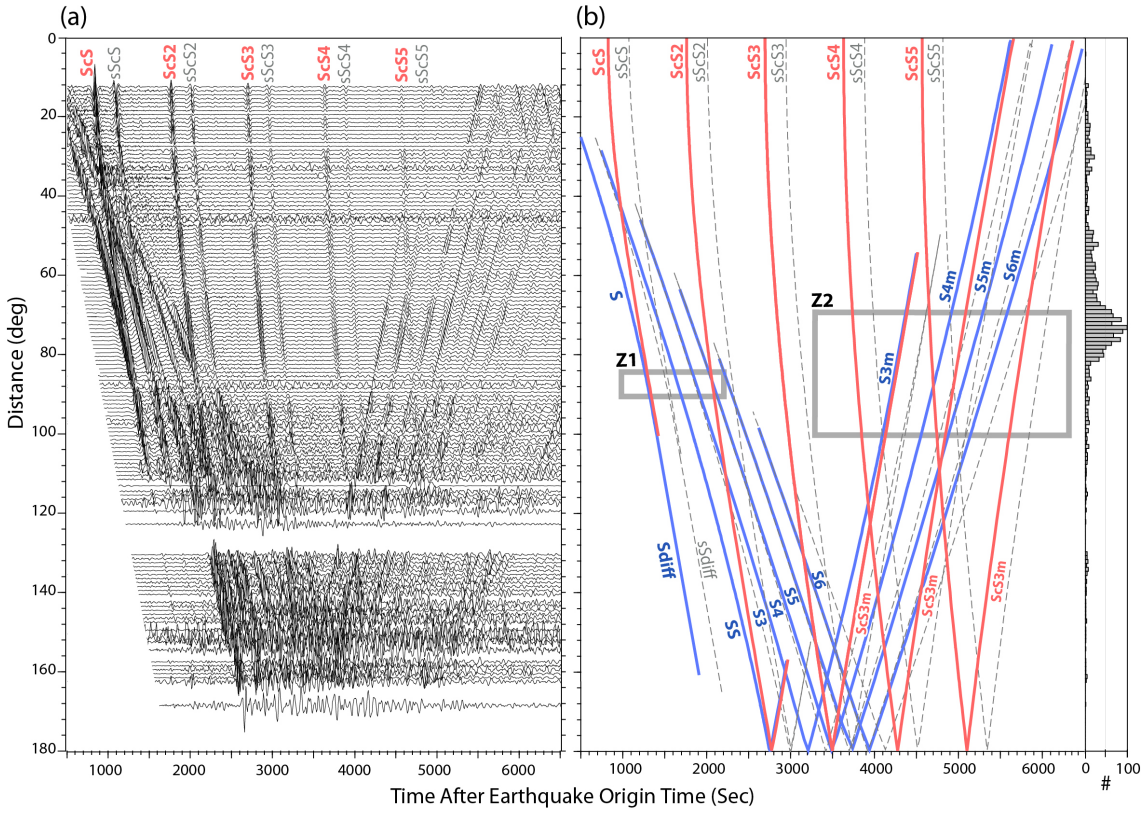

Figure 4.1. (a) Record section distance profile of transverse component velocity recordings for an event on May 24, 2013 (latitude: 54.87, longitude: 153.28, depth: 608.9 km, Magnitude: 8.3 Mw). Seismograms were low-pas filtered at 20 sec, then linearly stacked in 1 degree distance bins. (b) Travel time curves for the principle phases of interest are shown, and include direct S, Sdiff, and multi-bounce S waves (minor and major arc) as blue lines. ScS and higher multiple bounces are also shown (orange lines). Depth phases are shown as gray dashed lines. Any phase name with an "m" at the end is a major arc wave. The two gray boxes (marked Z1 and Z2) are two zoomed in time-distance regions, which are shown in Figure 2. The number of seismograms in each stack shown in panel (a) are presented on the right of panel (b) as a histogram. The most populated stacks have up to 100 seismograms.

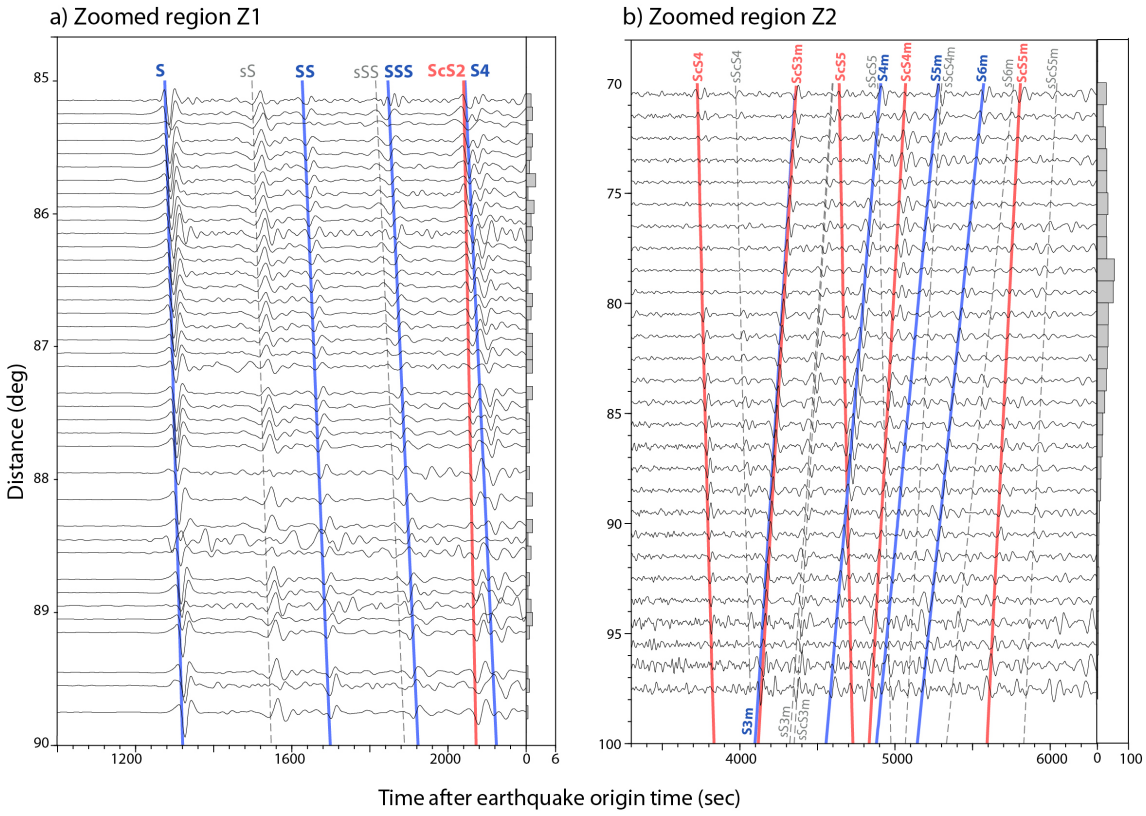

Figure 4.2. Zoomed in record sections for same event of Figure 4.1 and regions demarked in Figure 1b. As in Figure 1, records are stacks of small distance ranges. Tangential component velocity records are shown. (a) Zoomed area Z1 showing stacked records for phases S, SS, S3, ScS2 and S4. Dashed lines correspond to depth phases (colors are as in Figure 4.1). The distance increment used in stacking is 0.1 degrees, and number of records in each stack is shown in the histogram to the right of the record section. (b) Zoomed area Z2 showing stacked records for phases S3m, S4m, S5m, S6m, ScS4, ScS5, ScS3m, ScS4m, ScS5m, and depth phases. A larger time window is shown than for panel (a), and the distance increment used in the stacking is 1 degree. All else is as in panel (a).

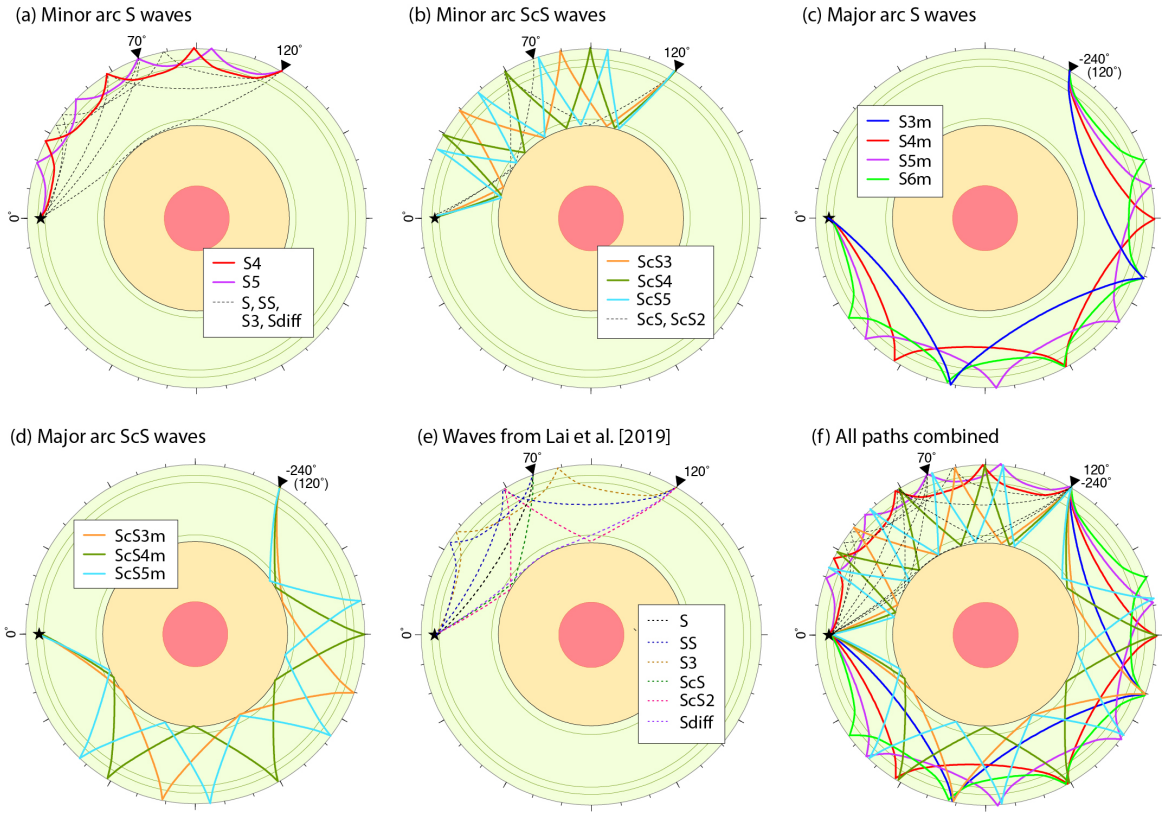

Figure 4.3. Cross-sections showing ray paths of a 500 km deep earthquake for (a) wave paths of minor arc multi-bounce S waves (S4 and S5, at 120 degrees) and paths of S, SS, S3, and Sdiff studied in Lai19 at 70 and 120 degrees; (b) as in (a), but minor arc multi-bounce ScS waves (ScS3, ScS4, and ScS5 of this study, ScS and ScS2 of LaI19); (c) major arc multi-bounce S waves (S3m, S4m, S5m, S6m at 120 degrees minor arc distance, -240 degrees major arc distance); (d) as in (c), but major arc multi-bounce ScS waves (ScS3m, ScS4m, and ScS5m); (e) the six seismic waves studied in Lai19; (f) all of the paths of (a) through (e) combined, with all phases pursued in this study given in color (the gray dashed lines are the 6 phases of Lai19).

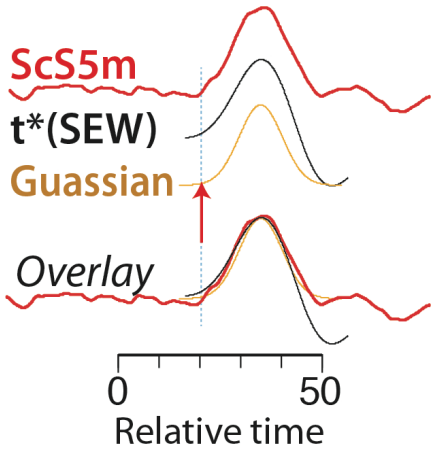

Figure 4.4. Example comparison of a major arc ScS5 virtual station stack (ScS5m, red traces) with an SEW that has been broadened by convolution with a t* operator to best fit the ScS5m stack. The Gaussian function that best-fits the broadened SEW is also shown (orange trace) along with a red arrow which indicates the automatically assigned onset time. The traces are overlayed to show the wave shape comparisons.

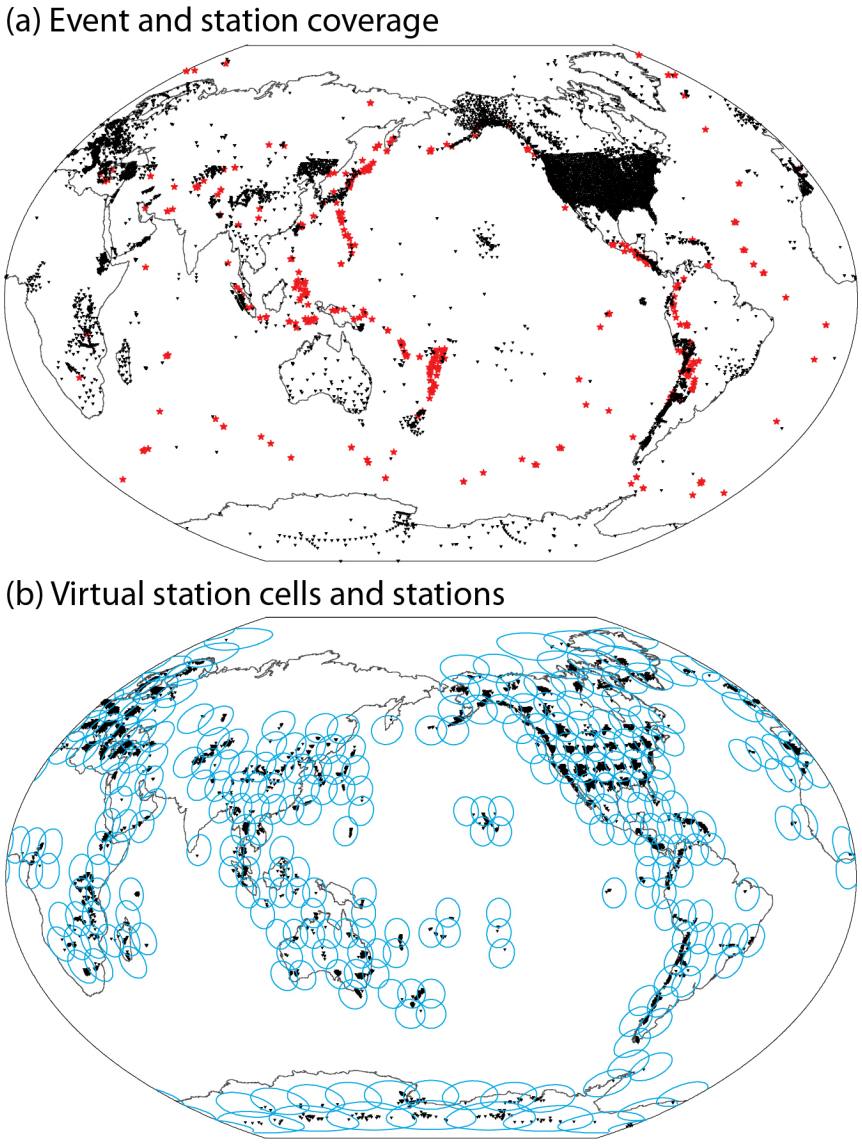

Figure 4.5. (a) Geographical locations of 360 earthquakes (red stars) and 8407 seismographic stations (black triangles) used in this study to explore virtual station construction. This is the same dataset as Lai19. (b) Locations of 289 initial virtual station grid cells (blue circles), and roughly 247K distinct relocated virtual station locations (black triangles).

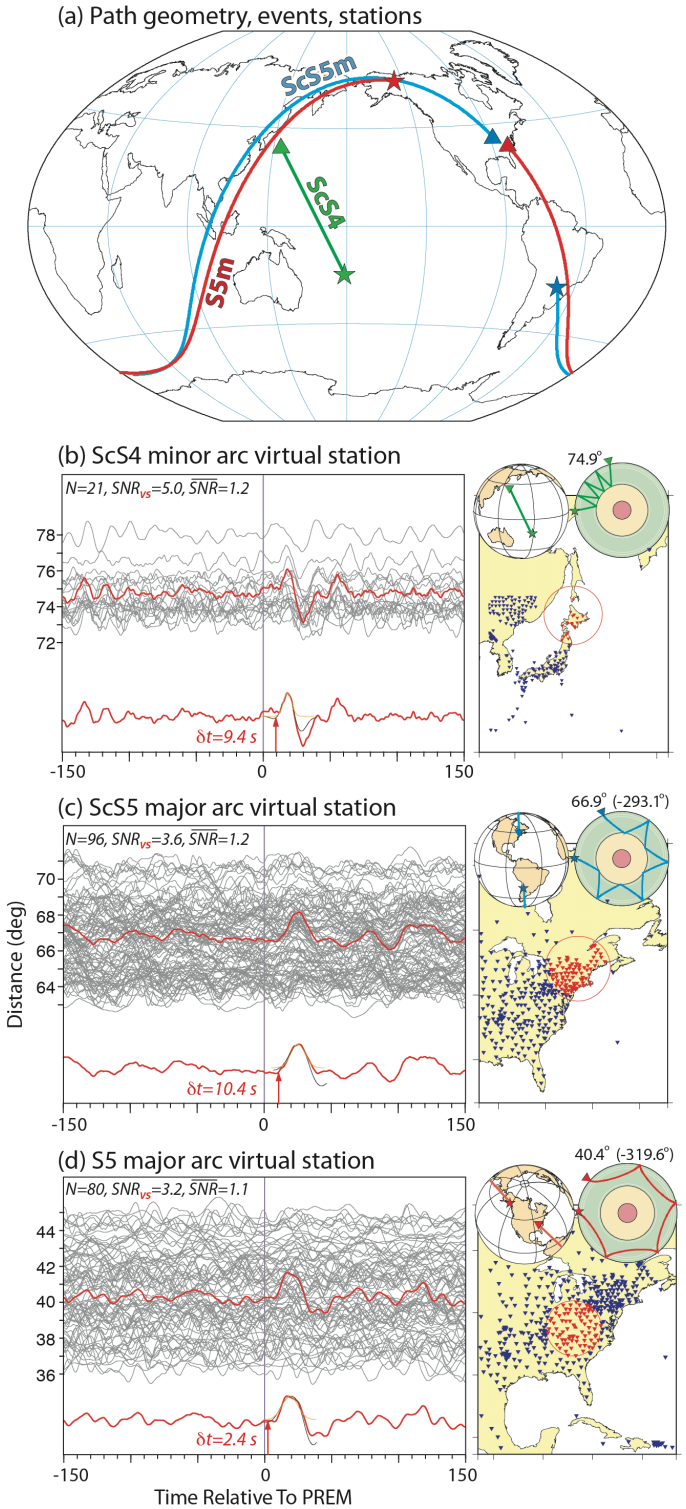

Figure 4.6. Virtual Stacking Examples. (a) Map showing great circle ray path geometries of examples shown in rest of the figure, with earthquake (stars) and virtual station bin

center (triangles) locations. (b) Record section of 21 transverse component velocity recordings of minor arc ScS4 (left, gray traces) for a virtual station bin in northern Japan (map to the right, red triangles are those of the traces on the left, blue triangles are the available stations for that event). Path geometry is also shown on the right (above the map). Virtual station stack is the red trace plotted at the distance of the virtual station bin center (74.9 degrees), and also plotted below with the best-fitting SEW and Gaussian function, along with the onset time predicted by the empirical wavelet method (red arrow), here, 9.4 sec after the PREM prediction. The number of stations in the stack (N), the SNR of the virtual station stack (SNR_{VS}), and the average SNR of the contributing records ($\overline{SNR}$) are printed in the upper left. (c) As with (b), except for ScS5m (ScS5 major arc). But minor and major arc distances are indicated above the ray path cross-section panel (upper right). (d) as with (c), except for S5m (S5 major arc).

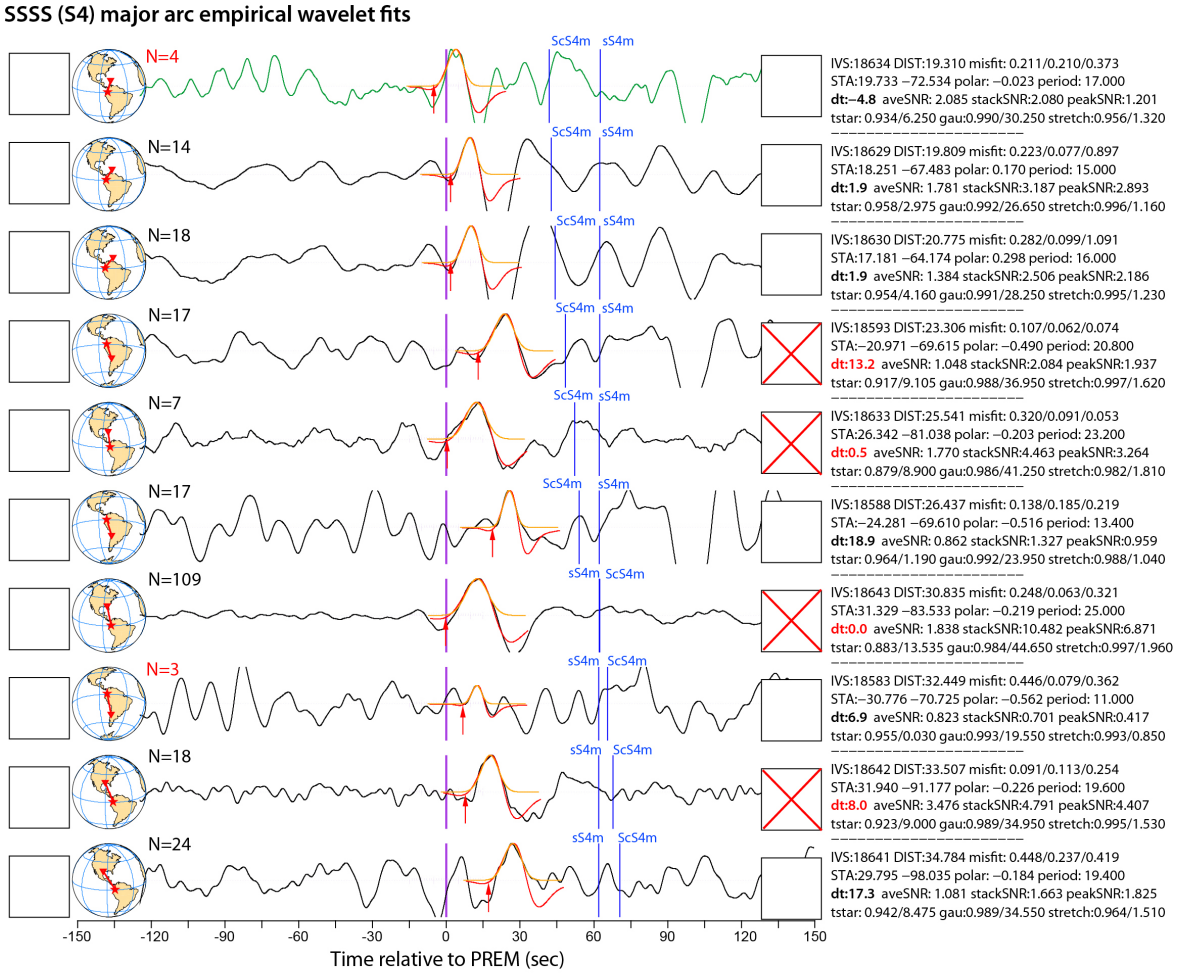

Figure 4.7. Partial page of an S4m virtual station PDF catalog plot. Virtual station stacks are black traces, adapted empirical wavelets (SEWs) and Gaussian wavelets (red and orange traces, respectively) are shown with onset time predictions (red arrows) as determined by the SEW fitting algorithm. All traces are plotted relative to the PREM-predicted time (purple lines at time=0, the predicted arrival time of S4m). Predictions for two expected arrivals (the depth phase sS4m, and ScS4m) are notated with vertical blue lines and text labels. Other features, from left to right, include: user text input box, reference globe (showing earthquake (red star), station (red triangle), the minor arc great circle path (red line)), the number of seismograms (N) used in the virtual station stack

(which is red when N<5), the retain/reject box (red X indicates record is retained), and a text block with information about the virtual station: ID#, epicentral distance (DIST), pre/signal/post misfit, station (STA) latitude and longitude, predicted radiation pattern amplitude (polar), estimated period, travel time anomaly relative to PREM (dt), $\overline{\mathbf{SNR}}$ (aveSNR), SNR_{VS} (stackSNR), SNR_{peak_trough} (peakSNR), and information about the t* operator, Gaussian factor, and stretching values.

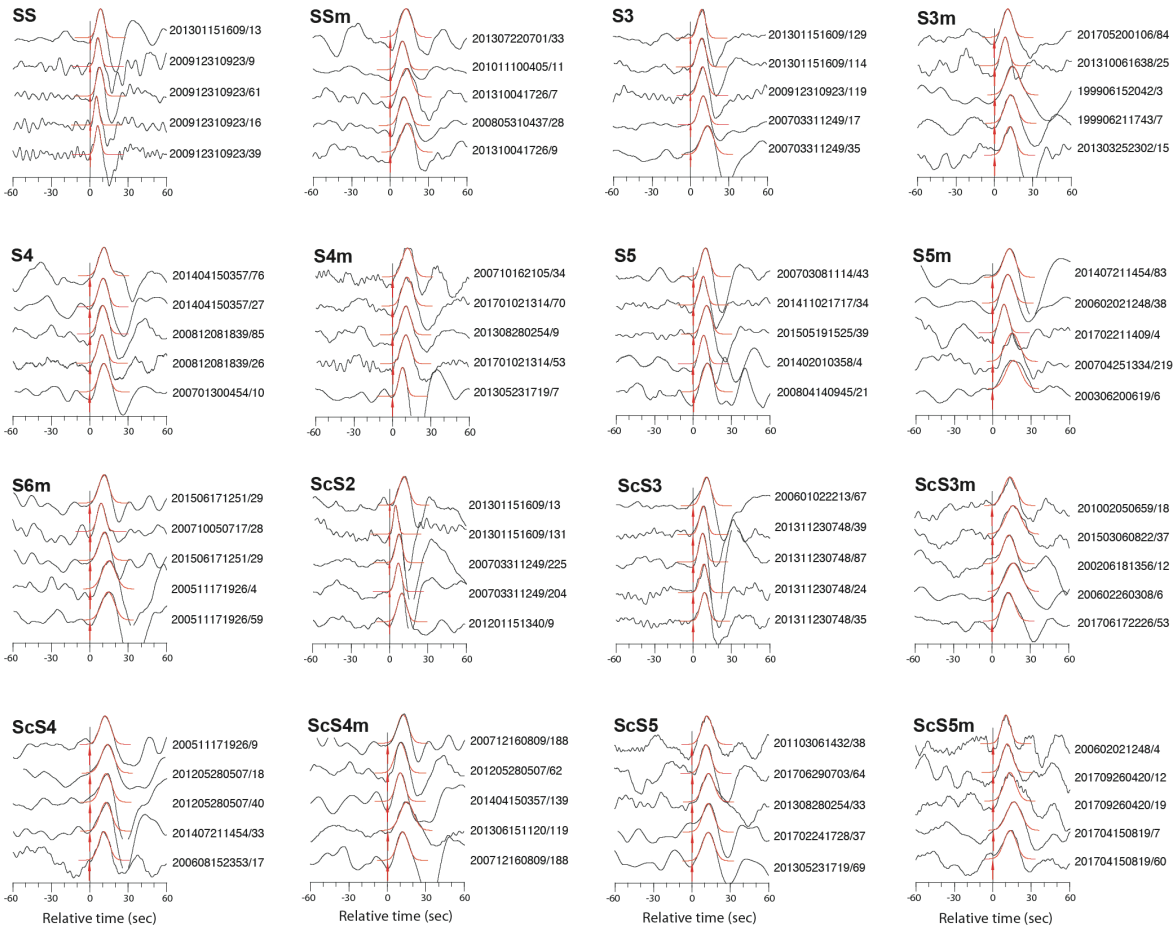

Figure 4.8. Five virtual station seismogram examples for the 16 multi-bounce phase types studied here (black traces) plotted with the best-fit stretched empirical wavelet (SEW, red traces) and the onset time determined from the Gaussian function that best-fits the SEW (red arrows). Event information and the number of seismograms used in the virtual station stack are to the right of each trace, as yyyymmddhhmm/N (yyyy=year, mm=month, dd=day, hh=hour, mm=minute, N=# records). Records are all lined up on zero time for ease of viewing, and all maximum amplitudes are normalized to unity.

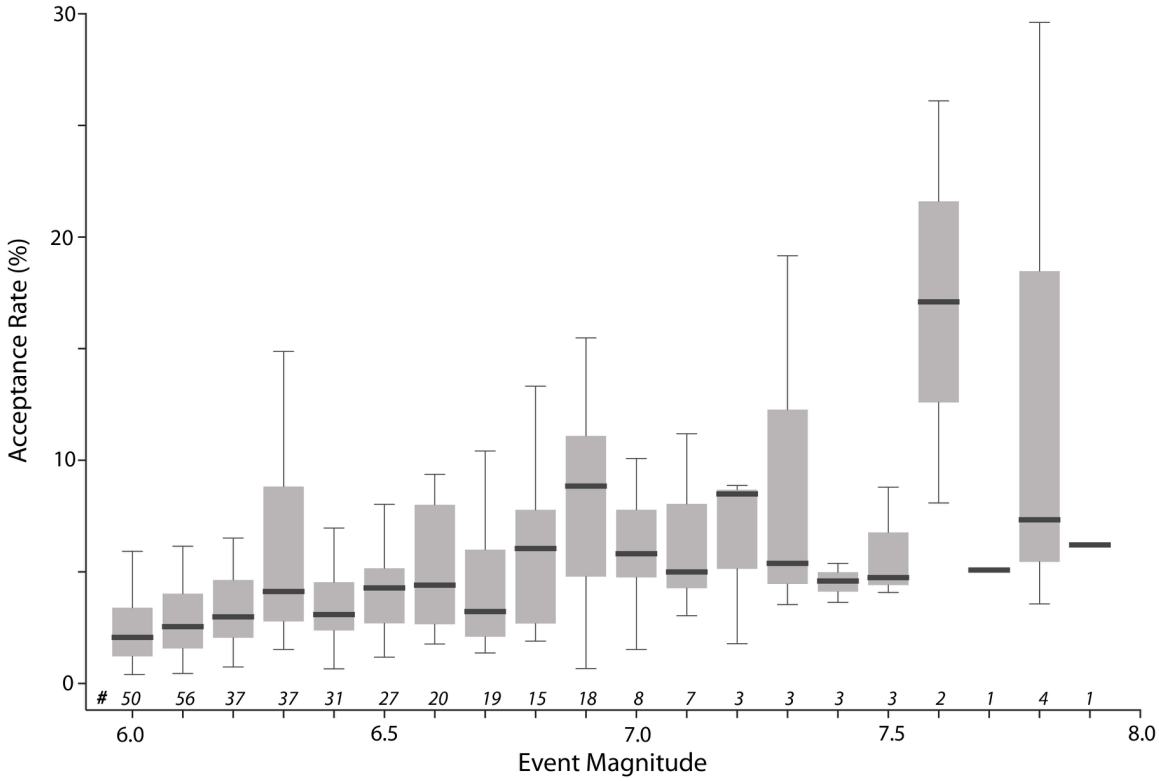

Figure 4.9. Virtual station measurement acceptance rate averaged in earthquake magnitude bins, for all measured phases of Table 4.1. Thick horizontal bar in the gray shaded box represents the median value for each magnitude bin population. The gray boxes show the range between the 25th and 75th quantile percentage ranges of the population (Q1 and Q3, respectively). The vertical bars outside of the gray boxes extend upwards to the largest value in the population within the maximal (Q3+1.5(Q3-Q1)) range and downwards to the smallest value in the minimal (Q1 – 1.5(Q3-Q1)) range. The italicized numbers above the magnitude axis correspond to the number of earthquakes used in the acceptance rate averages.

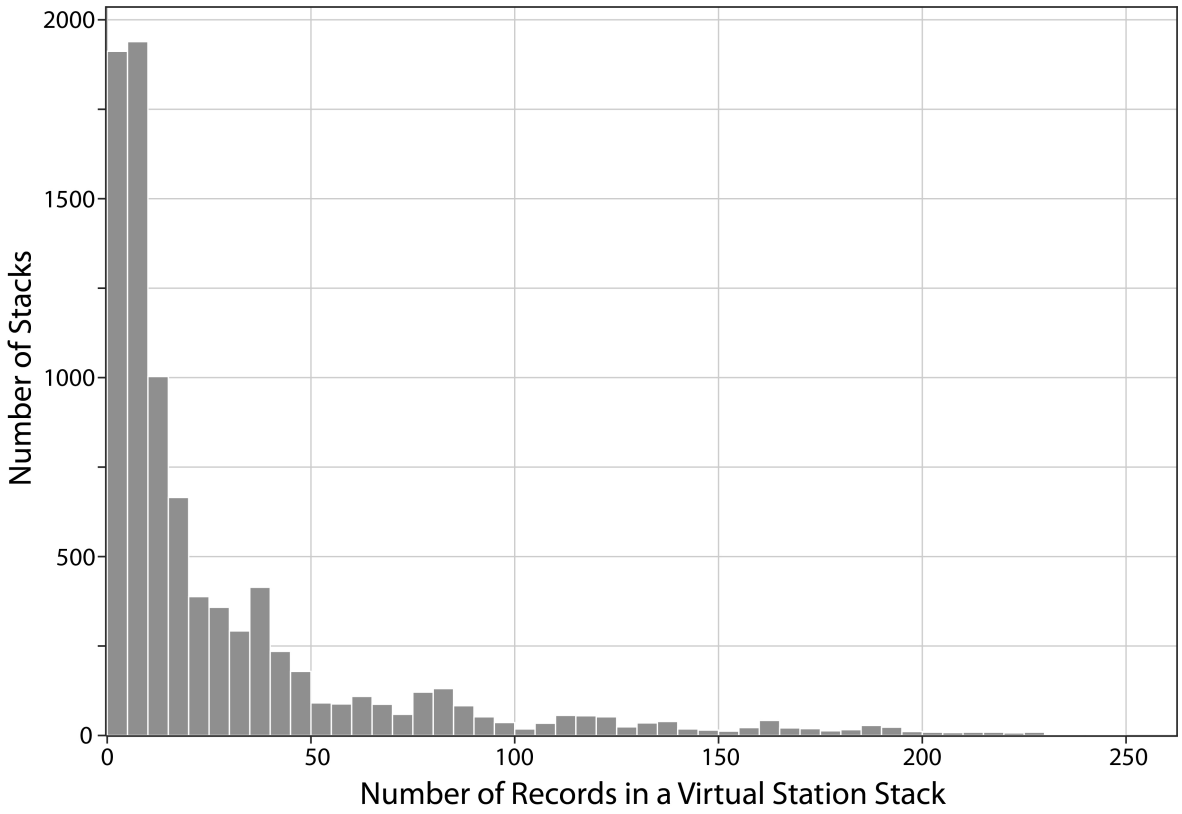

Figure 4.10. Frequency histogram showing the number of virtual station stacks for different numbers of contributing records. Over half of the population is in the first 3 bars (between 3 and 30 records). See text for more information.

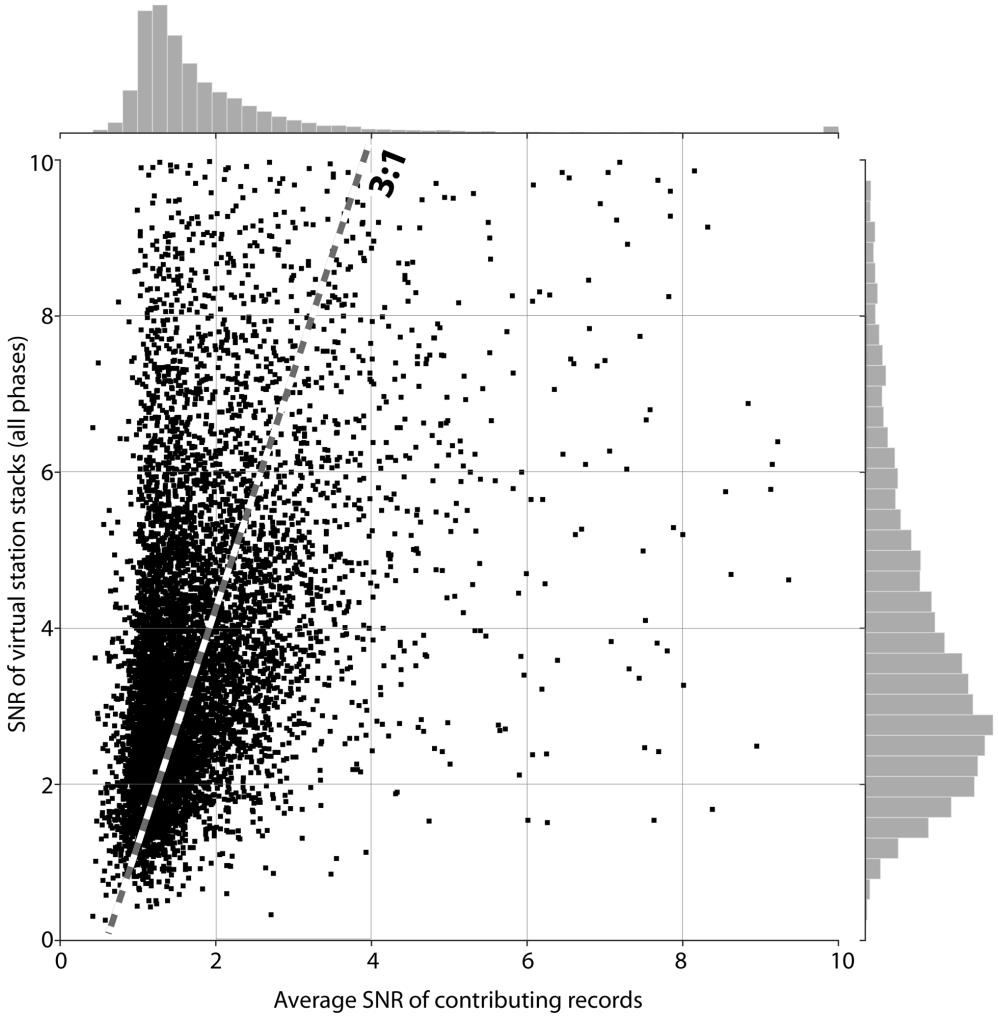

Figure 4.11. Comparison of virtual station stack SNRs to the average SNR of the contributing records for each stack. Included data virtual station stacks have a minimum of 3 contributing records. A line with 3-to-1 slope compares well with the data points. The histogram at the top of the figure corresponds to the number of virtual stacks as a function of SNR measurement; the histogram to the right corresponds to the number of record SNR averages taken, as a function of their average SNR.

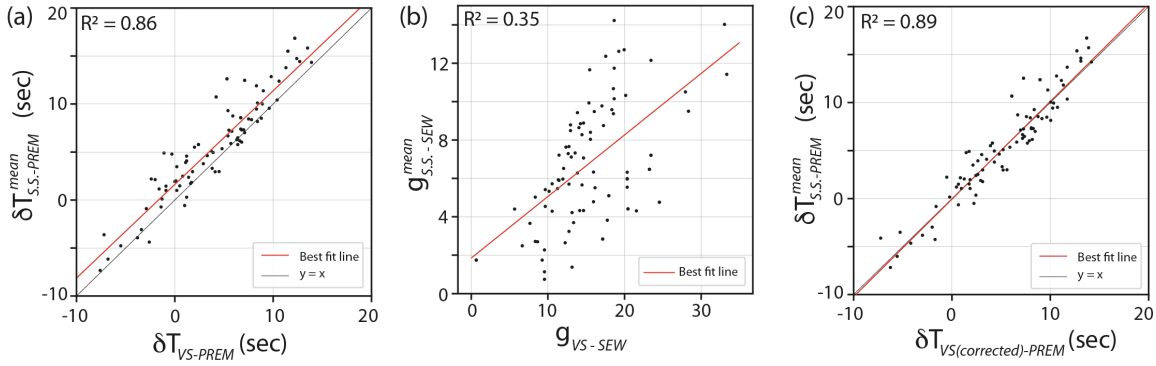

Figure 4.12. (a) For virtual station stacks containing single seismogram measurements, the average travel time anomaly (with respect to PREM), $\delta T^{mean}_{S.S.\text{-}PREM}$, is compared to the travel time anomaly of the virtual station stacks they contributed to, $\delta T_{VS\text{-}PREM}$. A best-fit line and R^2 value are also shown. (b) The average of the best-fit Gaussians of the single seismograms from (a), $g^{mean}_{S.S.\text{-}SEW}$, compared to the best fit Gaussian of the virtual station stack, $g_{VS\text{-}SEW}$, where both were referenced to the Gaussian of the SEW for their corresponding events. A best-fit trend line and R^2 value are also shown. (c) After the Gaussian of the virtual station stack is corrected using the trend line in (b), an onset time is determined, and compared to the average of the onset times of the contributing single seismogram records. The updated trend line has an R value of 0.89, and agrees well with the y=x line.

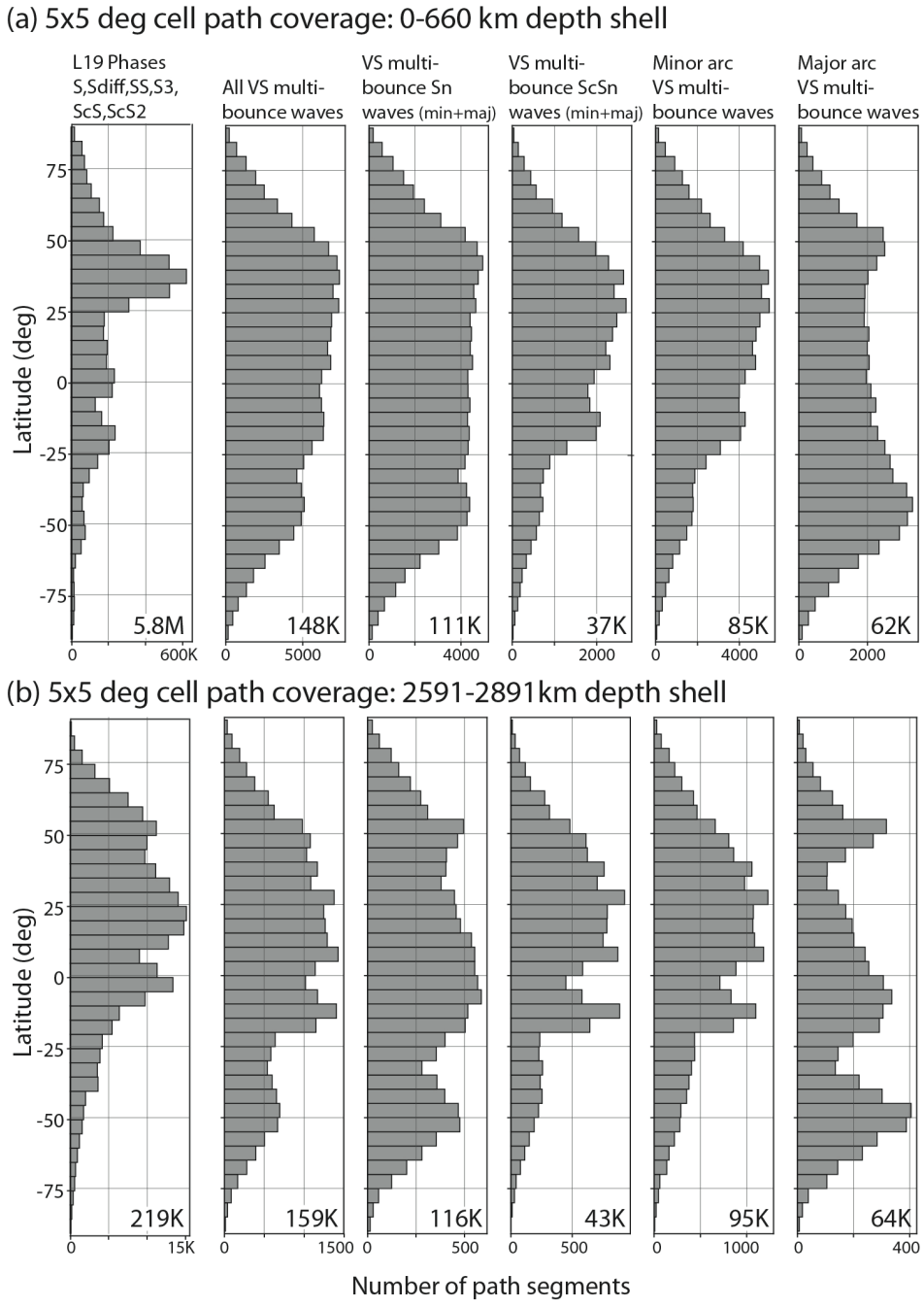

Figure 4.13. Ray path coverage as a function of latitude, where the number of rays were counted in all 5x5 degree cells, for a variety of different data groupings, for (a) the entire upper mantle (between 0 and 660 km), and (b) the lowermost 300 km of the mantle. The total number of path segments is written in the lower right of each panel. The single

seismogram data set of Lai19 (for S, Sdiff, SS, S3, ScS, and ScS2) is shown in the leftmost column. The second column presents coverage of all virtual station stacks measured in this paper. The third column shows the coverage of all major and minor arc multi-bounce Sn waves (n=2-6). Column 4 is the same as column 3, except for ScSn (n=2-5). Column 5 presents coverage for all minor arc phases and column 6 presents coverage for all major arc phases. See text for more details.

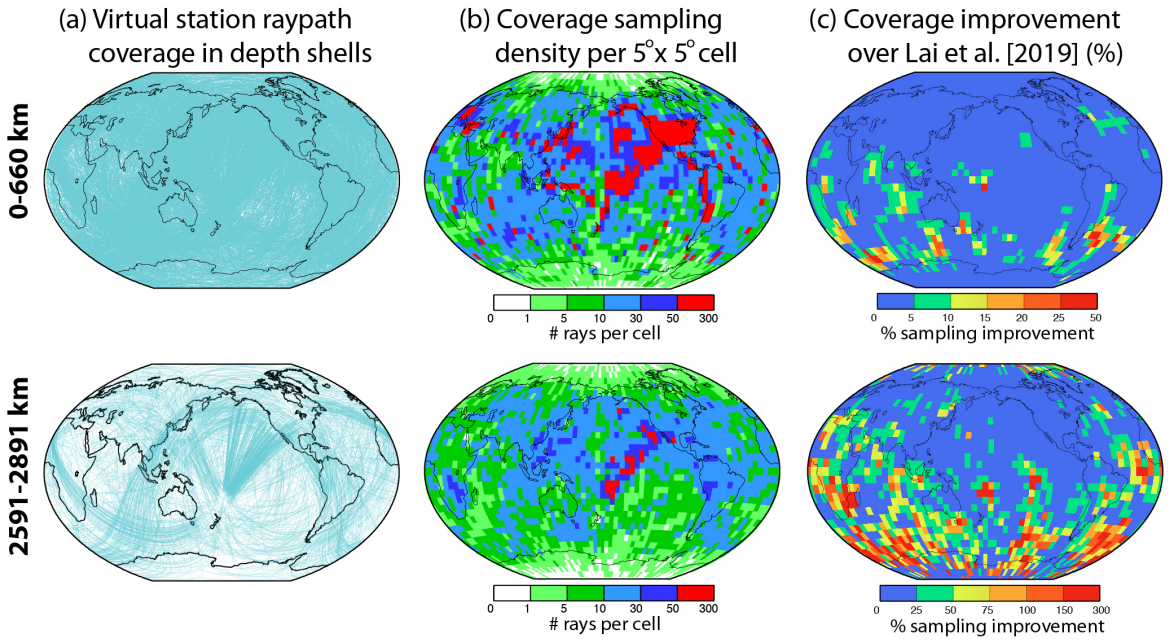

Figure 4.14. (a) Ray paths (blue lines) of all virtual station stacks measured in this study, for the upper mantle (top row) and deepest 300 km of the mantle (bottom row). (b) Coverage sampling density in 5°x5° cells, in the upper and lowermost mantle, for the ray paths in (a). Scale bars present the number of rays counted in each cell. (c) Sampling density coverage improvement over Lai19 (in %) by taking the panels in (b) and dividing by the same for the ray paths of Lai19. Thus 100 represents the same number of rays in each dataset. The warmer colors (strongest sampling improvement) are mostly present in the southern hemisphere, especially for the deepest mantle.

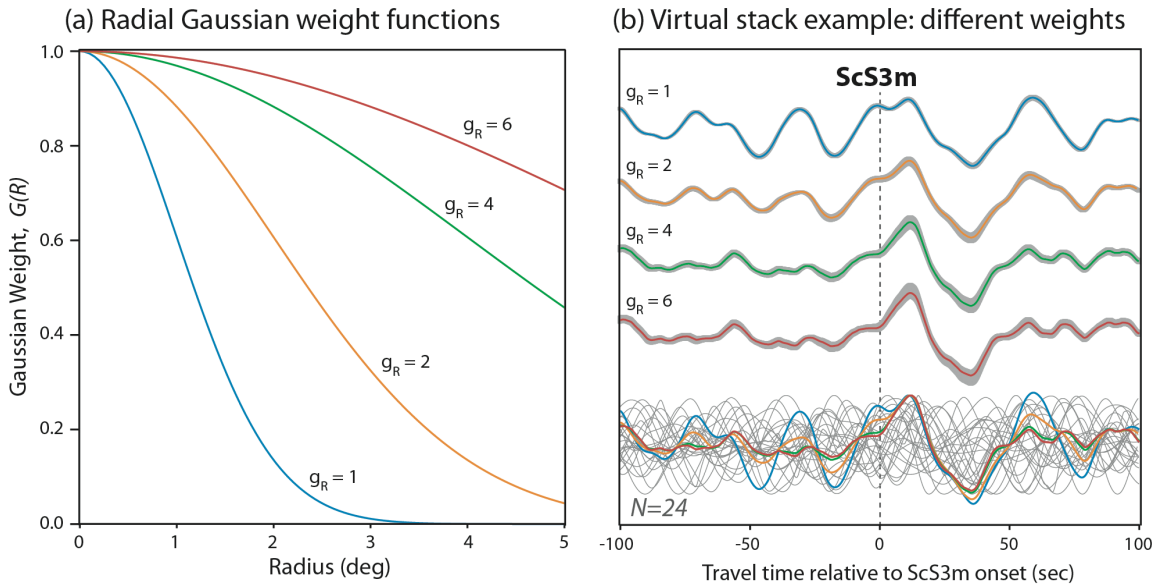

Figure 4.15. (a) Relationship between distance from grid center and Gaussian weight *G(R)* for different Gaussian factors g_R. This study used g_R=4. (b) Virtual station stack example for the major arc multi-bounce wave ScS3m. Virtual station stacks are shown for the same g_R of panel (a), along with $\pm$1 standard deviation (with amplitudes normalized to the maximum energy in the time window). The bottom displays the 24 records used to make the stack, along with the four virtual station stacks overlaid on them (the virtual station stacks are normalized to the amplitude to the of the estimated ScS3m peak).

CHAPTER 5

5 LAYER STRIPPING ITERATIVE FORWARD TOMOGRAPHY

Abstract

We present an iterative forward update method that iteratively updates a starting model with the goal to reduce the residual travel time variance. We utilize the seismic measurements documented from previous experiments in Chapter 2 and Chapter 4 and developed a novel layer stripping iterative forward tomography method to construct a new 3D seismic velocity model. This method updates a starting model iteratively from the surface to the core mantle boundary. The iterative forward approach is designed to restore the amplitude and pattern of seismic shear velocity that is usually heavily normalized and smoothed in seismic tomography inversion. Combined with the high-quality seismic dataset that we have documented, including the quality measurements used as weights, updated seismic models have been constructed with different starting tomography models. We compare the difference in both the starting model and updated models and we show more consistent features in both pattern and scale in the updated models. The consistency in updated models also shows independency from starting models and provide confidence that iterative forward tomography is updating starting models into a converged solution. Additionally, strong velocity perturbation is observed within LLSVPs, similar to the pattern predicted from geodynamic models, suggesting the thermochemical nature of LLSVPs. We have further identified several locations with low velocity anomalies in the updated models. The consistency between our updated models and with results from forward and inverse studies on the same region suggests higher resolving power comparing to the starting tomography models.

5.1 Introduction

Through inversion of observed seismic measurements, a 3D velocity model of the mantle can be constructed, which is termed 'seismic tomography'. Global seismic tomography aims at explaining large volumes of global seismic data and resolving whole mantle structure. Combined with geodynamic simulations and mineral physics experiments, seismic tomography serves as a powerful tool to study the structure and dynamics of Earth`s interior. Different from medical tomography, the ability to resolve the 3D structure is limited by restricted earthquake and station geometry, and also the ability to accurately model the effects of wave propagation in the real Earth [*Romanowicz*, 1991].With the expansion and deployment of new seismic networks, especially EarthScope's USArray (http://www.usarray.org, https://earthquake.usgs.gov), large volumes of high quality seismic data are available. Combined with high performance computing, more sophisticated methods such as full waveform tomography are used [*French and Romanowicz*, 2015; *Durand et al.*, 2017]. Our understanding of mantle structure has been greatly improved with the recent development of seismic tomography [*Becker and Boschi*, 2002; *Grand*, 2002; *Panning and Romanowicz*, 2006; *Houser et al.*, 2008; *Kustowski et al.*, 2008; *Li et al.*, 2008; *Ritsema et al.*, 2011; *Auer et al.*, 2014; *Moulik and Ekström*, 2014; *French and Romanowicz*, 2015; *Koelemeijer et al.*, 2016; *Durand et al.*, 2017]..

Results from different groups have shown the bottom 1000 km of the mantle is dominated by two large low shear velocity provinces (LLSVPs), one beneath Africa and the other beneath the Pacific Ocean and Southern Indian Ocean. Several studies have been conducted to compare the results from different seismic tomography models [*Lekic*

et al., 2012; *Garnero et al.*, 2016], emphasizing the robustness of very large-scale features identified in different tomography models and also pointing out the disagreement of small to intermediate-scale structures, especially in the lower mantle [*Becker and Boschi*, 2002; *Garnero et al.*, 2016]. The disagreement between different seismic tomography models could be caused by many factors, including the different data types and data quality used, the simplifications and assumptions of the theory behind the inversion technique, the parameterization of different models and the degrees of normalization and smoothing implemented during the inversion.

By taking a closer look at the intermediate-scale structure in the lower mantle, many interesting features are observed but are not consistently resolved in global seismic tomography, such as the low shear wave velocity anomaly beneath northern Pacific Ocean [*Suzuki et al.*, 2016] and the Perm Anomaly (low shear wave velocity anomaly) [*Lekic et al.*, 2012] beneath Perm, Russia. The smaller scale interior structure of LLSVPs also shows strong variations in S-wave velocity but is not well documented [*Garnero et al.*, 2016]. Better constrains on these intermediate structures would address some of the key questions, such as the thermal or compositional nature of these structures, the mantle dynamic patterns around these structures.

To address these issues, we developed an iterative forward update approach to provide better constraints on mantle structures, especially in the lower mantle. This method utilizes a high-quality seismic travel time dataset that is constructed from our previous work [*Lai et al.*, 2019]. Then we use this dataset to iteratively update a starting seismic tomography model, with the goal to map the residual that could not be explained by the starting model into the updated model. We choose to use the iterative forward

update approach to avoid some of the shortcomings of the inversion approach. Seismic inversion usually involves heavy regularization and smoothing. Due to the limited geometry of seismic data available, the inversion problem become an underestimated problem at locations where there is not enough data coverage. The biased sampling of seismic data usually causes instabilities in the constructed model, regularization (or damping) is introduced to smooth the model and suppress the variance, which often gives a more desirable result. To address this issue, most seismic tomography models use strong smoothing in data kernels to achieve a "complete" model. Global seismic tomography models aim to achieve a balance between model variance and model smoothness, which causes the resolved amplitude of the seismic velocity to be much lower than results from local forward studies [*Wang*, 2004; *He et al.*, 2006; *Sun et al.*, 2009].

The forward iterative approach allows us to map the high-quality seismic travel time dataset residual into the path of each individual data record. This approach allows us to control where to map the anomaly based on prior information. The consistency between different tomography models provides confidence on large-scale structures. And using a seismic tomography as our starting model drastically decreases the iterations needed for our model to converge.

We introduce the data used in this study in Section 5.2. The implementation of iterative forward approach is discussed in Section 5.3 and the constructed final models from different starting models are discussed in Section 5.4.

5.2 Data Set

5.2.1 Travel Time Dataset

The dataset used in this study is constructed using the Empirical Wavelet construction method, which systematically and objectively documents shear wave travel time. The phases and number of data entries in the dataset is listed in Table 5.1. As discussed in our previous work [*Lai et al.*, 2019], empirical wavelet are stacked iteratively from the data as the reference waveform on an event-by-event basis. The modified empirical wavelet is then cross-correlated with each record and the arrival time is derived through a modified Gaussian function that best matches the modified empirical wavelet. This approach provides a way to document the arrival time of each seismic record in a systematically and objective manner and it also allows the whole process to be automated, and thus is applicable to global seismic datasets. Also, this method systematically accounts for the waveform broadening effect caused by either structure or attenuation, which we believe provides more accurate measurement of seismic travel times. In the end we have constructed a travel time dataset for S, SS, SSS, Sdiff, ScS and ScSScS (referred as the 6 basic phases) with a total of ~250K data entries.

From the coverage of this dataset, the coverage in the Northern hemisphere is significantly better than the Southern hemisphere, with 5 times more sampling in the Northern hemisphere (through comparing hit count in northern and southern hemispheres). Most of the events that we use to construct the 6 basic phases dataset are deep events that are concentrated at subduction zones and plate boundaries, such as the Fiji-Tonga region, the west coast of South America and the western margin of the Pacific Ocean. In our work, to diversify the distribution of event locations, we also introduced

shallow events, which are more scattered than deep events and thus more likely to provide unique source location. Due to the sheer number of shallow events available, we have developed a ranking algorithm that is based on the distance between shallow events and the existing deep events in our dataset. We adopted this algorithm to prioritize events with unique locations. Even with the optimized event distribution, the distribution of earthquakes and seismic stations is still very limited, combined with the limited phases we use, the sampling bias towards Northern hemisphere is as expected.

During the empirical wavelet construction, 80% of the total number of seismic records are classified as noisy records due to the low signal-to-noise ratio and low quality, especially for seismic networks outside of north America. Utilizing the vast amount of noisy data potentially can provide us much more usable data and more importantly, could improve our limited existing sampling coverage. We have designed a virtual stacking method to stack neighboring stations to improve signal-to-noise ratio to thus make use of the seismic records that are noisy when viewed individually. We have implemented the virtual stacking method on 360 earthquakes and have retrieved about 5000 virtual stations for the 6-basic phases listed in Table 5.1.

Another important attribute of this dataset is that it includes a weighting factor for each individual data entry. The weight is defined by a combination of factors that describe different characteristics of the record, including the signal-to-noise ratio (SNR), the cross-correlation coefficients (CCC) and the misfit measurements. As SNR being one of the most important measure of the quality of the data, we have defined different types of signal-to-noise ratio, including SNR based on average amplitude between the signal window (one period window centered around the phase) and noise window (80 seconds

in front of the phase), SNR based on maximum peak-to-trough amplitude within one period for the signal as well as for the noise window, SNR based on maximum and minimum trough amplitude in the signal and noise windows. The cross-correlation coefficients describe the waveform similarity between the record and the adapted Empirical Wavelet with values ranging between -1.0 and 1.0. The waveform misfit measures the degree of waveform difference between the record and the adapted Empirical Wavelet. It is defined as the average amplitude difference measured in the one period window centered around the phase. The weight is then defined by a combination of the three measurements and it describes the relative quality of our data, which allows us to put more confidence on high quality data and reduce weighting of those with low weight.

5.2.2 Correction Terms

To resolve the seismic velocity in the mantle, we have to first remove the effect of the crust. We have used the CRUST 1.0 model [*Artemieva and Mooney*, 2001] to account for the isotropic V_s and to correct for travel time in the crust.

Earthquake hypocenter mislocations are very common and can cause significant bias in calculated residual travel times. The generic forward model for event relocation is described in Spakman & Nolet (1988). Here we do a linear inversion to correct for errors in hypocenter origin time, depth, latitude and longitude. We use $\boldsymbol{\delta h}$ to represent the hypocenter origin time change (dt_0) and the hypocenter location change vector (dz the change in depth, $d\theta$ the change in longitude in radians and $d\phi$ the change in colatitude in radians):

$$\mathbf{\delta h} = \begin{bmatrix} dt_0 \\ dz \\ d\theta \\ d\phi \end{bmatrix} \tag{5.1}$$

We format the linear system as:

$$\delta \mathrm{T} = \mathbf{D}\delta \mathrm{h} \tag{5.2}$$

where δT is the perturbation of travel time for all phases of interest and **D** is the sensitivity matrix that describes the relationship between the change in travel time and change in earthquake location:

$$\mathbf{D} = \begin{bmatrix} 1 & \frac{\partial T}{\partial z_1} & \frac{\partial T}{\partial \theta_1} & \frac{\partial T}{\partial \phi_1} \\ \cdot\cdot & \cdot\cdot & \cdot\cdot & \cdot\cdot \\ 1 & \frac{\partial T}{\partial z_m} & \frac{\partial T}{\partial \theta_m} & \frac{\partial T}{\partial \phi_m} \end{bmatrix} \tag{5.3}$$

Combined with model and data weighting matrix, we solve δh with matrix inversion:

$$\delta \mathrm{h} = [D^T W_d \mathrm{D} + \epsilon^2 W_m]^{-1} D^T W_d \delta \mathrm{T} \tag{5.4}$$

where the model weighting matrix is:

$$W_m = \begin{bmatrix} 1/1000 & & & \\ & 1/6371 & & \\ & & 1 & \\ & & & sin\theta \end{bmatrix} \tag{5.5}$$

The weighting value is selected to normalize the model parameters so that they have the same effect when measuring model length. The data matrix is:

$$W_d = \mathrm{diag}(\mathbf{W}) \tag{5.6}$$

$$W_i = t_i * b_i * m_i \tag{5.7}$$

where t_i is the amplitude weight that is associated with the travel time residual amplitude following the rules below:

$$
\begin{aligned}
0 < |residual| < 3s, t_i = 1 \\
3 < |residual| < 6s, t_i = 0.8 \\
6 < |residual| < 9s, t_i = 0.6 \\
9 < |residual| < 12s, t_i = 0.4 \\
12 < |residual| < 15s, t_i = 0.2 \\
15\text{s} < |residual|, t_w = 0.1
\end{aligned}
\tag{5.8}
$$

b_i is the weight that is associated with each phase, where:

$$
Phase = S, b_i = 1.0, Phase \neq S, b_i = 0.6 \tag{5.9}
$$

and m_i is the weight inherited from the individual record from the Empirical Wavelet construction. The data weighting matrix is designed to build confidence in records that we believe have high quality and reduce weighting for those that have low quality. We solve the linear inversion to suppress the variance that we see in data.

The linear inversion is an overestimated problem and is similar to seismic tomography, there is tradeoff between model complexity and error, which is shown in Figure 5.1. As the model gets more complex, the error (travel time variance) decreases. We want a solution that minimizes the error but a simple model with lower variance is also desired. We select a model solution that is at the 'elbow' point, denoted by a red dot in Figure 5.1 as our final solution.

As the Earth is not a perfect sphere, seismic travel times need to be corrected for the effect of the ellipticity of the Earth. We use method introduced in [*Kennett and Gudmundsson*, 1996] to implement the ellipticity corrections.

5.2.3 Starting Tomography Model

The consistent large-scale structure we see in seismic tomography from different groups suggest robustness of these structure in the resolved model, even though the amplitude might be different from model to model. In our iterative updating scheme, using a seismic tomography result as a starting model acts as an effective way to reduce the number of iterations needed to resolve the final model. By using different starting models, we are also able to test the dependency and convergence of the final updated model.

We chose 5 recent tomography models as our starting models, including SAW24B16: [*Mégnin and Romanowicz*, 2000], GyPSuM-Shear wave model: [*Simmons et al.*, 2010], S40RTS: [*Ritsema et al.*, 2011], SEMUCB_WM1: [*French and Romanowicz*, 2015] and SP12RTS-Shear wave model[*Koelemeijer et al.*, 2016]. The 2800 km depth horizontal slices of the 5 models are shown in Figure 5.2. Consistent large-scale features, namely the two LLSVPs, one beneath Africa and the other one beneath the Pacific Ocean are observed across all models, while different models exhibit different patterns at intermediate to small scales, both within LLSVPs and outside of LLSVPs.

5.3 Iterative Forward Tomography method

In this section, we introduce the implementation details of iterative forward tomography. We first introduce the model parameterization and the layer stripping update scheme. The updating procedures within each iteration are explained in detail, including the weighting profiles used during the update and the smoothing parameters.

5.3.1 Model Parameterization

We divide Earth's whole mantle into cell blocks. For vertical parameterization, we divide the mantle into 100 km layers from the surface to the core mantle boundary, with exceptions only at the transition zones (namely 410km and 660km) and core-mantle boundary (2891 km). For horizontal parameterization, we use 2 degree by 2 degree blocks. Since the different starting tomography models used in this study come with distinctive parameterization, we converted the starting tomography models into the same format as our cell blocks through interpolation. Considering the high frequency band of the data used in this study (from 16 seconds to 100 seconds), we use infinite frequency ray kernels to represent the data kernel of each data record. The ray path of each individual record is generated with TauP toolkit [*Crotwell et al.*, 1999] using PREM (Preliminary Reference Earth Model, [*Dziewonski and Anderson*, 1981]).

5.3.2 Updating Procedures

From the iterative forward update perspective, to resolve the mantle structure of the lower mantle, we need to first resolve the mantle structure at shallower depths. Layer-Stripping tomography is a method that is commonly used in oil exploration which iteratively updates the model space in a top-to-bottom manner [Guillaume, 2012]. Since we focus on resolving the lower mantle structure, we adopt this top-to-bottom updating scheme and apply it in our iterative forward update approach. We first divide the model into two layers, layer I (from depth 0 to 2200 km) and layer II (from 2200 km to the core mantle boundary). The layer boundary is chosen to be 2200 m to preserve the lowermost ~700km as a separate layer.

With the layer stripping technique, we introduce a two-step updating scheme. We first use phases that have bottoming depths within layer I to update layer I, then in step two, we `freeze` layer I and only allow the velocity of cell blocks in layer II to vary. The details of the two steps updating scheme is presented below:

Step 1: Use travel time of S, SS, SSS and differential travel time pairs SS-S, SSS-S that have bottoming depth within layer I to update layer I. For the differential travel time pairs, the differential travel time of SS-S is allocated to path of the corresponding SS and SSS-S is allocated to the path of corresponding SSS.

Step 2: Use travel times of SS, SSS that have bottoming depths within layer II, together with travel times of ScS, ScSScS, Sdiff and differential travel time pairs ScS-S to update layer II. Similar to step 1, the differential travel time ScS-S is allocated to the path of ScS only.

The sequence update procedure is chosen to update the model in a top-to-bottom manner, similar to layer stripping tomography. Once the top layer is updated, we freeze that layer and do not allow further updates in that layer. Then we move to the next layer to continue the updating procedure. By freezing a layer and updating the subsequent layer, artificial velocity contrast might be introduced at the boundary between the two layers. To address this issue, we designed a smoothing kernel with gradual weight change at the layer boundary.

We have included wave pairs in our update scheme as well (if they exist), such as SS-S, SSS-S, ScS-S and ScSScS-SS pair. These wave pairs share similar earthquake and station geometries. By taking the travel time difference, we can remove the contribution cause by the same event and station region. During our update steps, the later phase in each phase pair is always updated before the phase pair. Therefore, we assume the later

phase path is already corrected for before we update for the phase pairs. We can then only assign the travel time difference on the path of the former phase (SS in SS-S pair, for example).

During each iteration, we follow the path of each data record to calculate the predicted travel time in each cell based on a starting model and then aggregate all incremental times to get the total predicted travel time. The total predicted travel time is then subtracted from the observed travel time in our dataset to derive the residual travel time. The residual travel time is used to construct an updated model for each iteration and in subsequent iterations the updated model from the previous iteration is used as new starting model. This process is implemented iteratively.

After we have calculated the travel time residual associated with each record, we assign the residual travel time along the path based on a combination of weights. Since we are using an iterative forward update scheme, it is important to incorporate prior information as weights to guide the direction of the update. By analyzing different global seismic tomography models, [*Becker and Boschi*, 2002] show that the root-mean-square heterogeneity of global seismic tomography models are much stronger in the upper and lowermost mantle than the middle mantle. The RMS heterogeneity of each seismic tomography model is then adopted in our method as a weight assigned to the travel time based on the depth of each given cell. The RMS heterogeneity weight is only applied to the first iteration of each updating step to avoid artifacts imposed by this rule. For subsequent iterations, we adopt weights similar to the RMS heterogeneity-based weight by using the ratio between the travel time anomaly and its corresponding path length.

This ratio is then averaged across each layer and represents the unit path travel time anomaly, which we use as proxy for RMS heterogeneity.

After assigning travel time residuals for all records, the cells in our model are filled with travel time residuals associated with small path segments contributed from different data records. We first convert the travel time residual into velocity perturbation using the equation below:

$$dVs_l^{delta} = \left(\frac{dl_l}{Vs_{i,j,k}^{PREM}\left(\frac{dl_l}{Vs_{i,j,k}^{PREM}\left(1+dVs_{i,j,k}^{prev_{tomo}}\right)}+dt_l^{residual}\right)} - 1 \right) * 100 - dVs_{i,j,k}^{prev_{tomo}} \tag{5.10}$$

where dvs_p is the velocity perturbation for the lth path segment for each data record and dl_l is its associated path length. $dt_l^{residual}$ is the travel time residual that is assigned to current path segment. $Vs_{i,j,k}^{PREM}$ is the transverse shear velocity predicted by PREM for the cell where the path is in and i, j, k represents the index corresponding to depth, latitude and longitude in the model. Similarly, $dVs_{i,j,k}^{prev_{tomo}}$ is the corresponding shear velocity retrieved from the updated model from previous iteration.

Within a given cell, the velocity perturbations contributed from different path segments are different due to many factors, such as the noise in the data, the travel time residual assigning kernel introduced in our updating procedures and the different geometry of each data record. To achieve a proper estimate of the velocity perturbation for each cell block, we calculate the weighted average of the velocity perturbation associated with each path segment. The weight we adopted here is based on the data

quality and azimuth of each data record. The following weighting scheme is implemented: a) Weight that is inherited from the data quality measurement from the Empirical Wavelet for each data record b) Weight associated with the azimuth of each data record. Balancing the velocity contribution from different azimuths is important, as for most cells in our model the azimuth distribution is heavily biased towards a certain geometry, such as the Fiji to North America geometry. We divide the azimuth distribution into 30 degree bins and each bin is assigned the same weight, each record is down-weighted by the total numbers of records in that same azimuth bin to balance the its contribution to the final velocity perturbation. c) Weight associated with different phases, which are listed in Table 5.2.

5.3.3 Smoothing

After converting the residual travel time to velocity perturbation for each cell, we implemented a weighted average of all contributing velocity perturbations. Due to the biased event-station geometry, each cell also has biased azimuthal sampling, causing certain azimuths to be much more heavily sampled than others. With experiments, we have realized that biased azimuthal coverage results in significant streaking in the updated final model, especially in the lowermost mantle. To improve the azimuthal coverage, neighboring azimuthal information is also utilized.

While we calculated the velocity perturbation through the weighted average of all contributing velocity perturbations within each cell, we further expand the range to include velocity perturbation contributions from neighboring cells as well. We define a radius of R (degrees) to search for neighboring cells and a Gaussian based radius weight is applied in the weighted smoothing to reduce weighting of cells that are far away. The

radius R that we adapt changes based on the azimuthal coverage available. We increment R from 2 degree to 6 degree, 10 degree and 14 degree. And at each incremental step, we search for the number of existing azimuthal bins (30 degree incremental azimuthal bins) where there is more than 1 sampling and if we have more than 5 unique azimuthal bins, we use that corresponding R value to select neighboring cells.

The smoothing scheme uses neighboring information to update the current cell velocity perturbation, it reduces the artifacts caused by model parameterization and therefore creates a smoother model that is desired. At locations where the sampling is sparse to even non-existent, smoothing with neighboring cells also helps us to achieve a complete model.

In Figure 5.3, we show the workflow of the iterative forward update, including the travel time residual calculation, residual travel time assignment with weights, residual travel time to velocity perturbation conversion and adaptive Gaussian smoothing.

5.3.4 Stopping Criteria

The goal of the iterative forward update approach is to update the 3-dimensional seismic velocity model to reduce the difference between travel time from model prediction and travel time observation in our dataset. We use the residual travel time standard deviation to describe the updated model's ability to explain the travel time in the dataset, which is defined as:

$$\sigma_{residual} = \sqrt{\frac{\sum(t^{data} - t^{model\ prediction})^2}{n}} \tag{5.11}$$

where t^{data} is the travel time in our dataset and $t^{model\ prediction}$ is the prediction predicted from data model of current iteration. n is the number of records that is used in each updating step.

During each iteration, we assign the residual travel time along the path of each data record and therefore our model`s ability to predict the travel time would improve as we do more iterations. We compare $\sigma_{residual}$ between two iterations, and if the standard deviation reduction is smaller than 3%, we would terminate the update in the current step and move to the next update step. This is introduced to avoid mapping unexplainable noise in our data into the model space and at the same time reduces computation cost.

5.3.5 Model Resolution

To understand the model resolution, we use a checkerboard to quantitively test the inaccuracies of the model. We use a checkerboard that has alternating high and low velocity anomalies with 30 degree distance alternating in both latitudinal and longitudinal direction. The amplitude is set to -3% and 3% relative to PREM model, respectively. In Figure 5.4a we show the input structure with color representing the velocity perturbation relative to PREM model in percentage. The output structure is plotted in Figure 5.4b for 600 km depth and Figure 5.4c for 2800 km depth.

The output model at 600 km depth shows very robust reconstruction of the input velocity patterns with minor smearing while the output model at 2800 depth is nosier and has a larger smearing effect. The discrepancy we see between the 600 km depth and 2800 depth is caused by the difference in azimuthal sampling in the dataset. For the 600km depth, phases like SS, SSS can travel very long distances up to a great arc distance of 180

degrees, which allows us to have event and station geometry that span across the whole globe, greatly enhancing the azimuthal coverage. In the lowermost mantle, the only data source comes from deep turning S waves, ScS, ScSScS and Sdiff, which all have limited sampling in the lowermost mantle. Deep turning S waves and ScS waves also have limited distance range (0 to 70 degrees for ScS and 30 degree to diffraction distance for S wave) which limit the event-station geometry that we are able to use. To accommodate this effect, we have used adaptive Gaussian radius-based weights to average neighbor cells to enhance the azimuthal sampling. Figure 5.5 shows the azimuthal coverage of our dataset, plotted at 600 km (top panel) and 2800 km (bottom) respectively. In the figure we show the azimuthal coverage of each cell every 10 degrees in both latitudinal and longitudinal direction, corresponding to each of the clustered arrows in the figure. For each cell, a total number of 12 azimuthal bins exist, and each corresponds to one arrow in the cluster. The arrow is only shown when there is sampling in the corresponding azimuthal bin.

5.4 Results and Discussion

We have implemented the iterative update scheme using S40RTS as starting model and the results are shown in Figure 5.6. Constant depth 'slices' of the model are plotted at 600 km, 1500 km and 2800 km depth. For each depth, the starting model is plotted at the left side, the updated final model is plotted at the right side and the difference is plotted in the middle.

We have further implemented this method with 5 different global seismic tomography models as the starting model, including SAW24B16 [*Mégnin and Romanowicz*, 2000], GyPSuM [*Simmons et al.*, 2010], S40RTS [*Ritsema et al.*, 2011],

SEWUCB_WM1 [*French and Romanowicz*, 2015], SP12RTS [*Koelemeijer et al.*, 2016]. All models are converted into our model parameter space and then used as the initial starting model in the iterative forward update. We show the final results with each starting model, where constant depth slices at 600 km depth have been plotted in Figure 5.7 and constant depth slices at 2800 km depth have been plotted in Figure 5.8. The starting model is shown as the left panel and the final updated model is show as the right panel. The difference is plotted as the middle panel.

By plotting different starting seismic tomography models with the same color scale, it is easier to identify the differences in both pattern and amplitude. After the iterative forward update, the updated models show more consistent features. It reduced the amplitude difference that we see in different starting models, especially for S40RTS and SP12RTS where the overall amplitude is enhanced. The updated models also show more consistent velocity structure patterns. The low shear velocity zone identified through forward modeling in Lekic et al. (2012) suggests a cylindrical structure with a diameter of 900 km beneath Perm, Russia. This structure is identified in their study with cluster analysis on 5 different tomography models, however the scale and amplitude as shown in Figure 5.8 are different. In the updated models, a more consistent structure is identified, with a much higher amplitude in velocity perturbation. By comparing the updated models, we have also identified a consistent low velocity zone beneath the east coast of North America outside of the Pacific LLSVP. This feature is consistent with the results in Nelson and Grand (2018), where a low shear velocity anomaly is modeled with a diameter of 350km that extends from the core-mantle boundary to the surficial position of the Yellowstone hotspot. At the northern side of the Pacific Ocean, a continuous low

velocity anomaly spans from the east side of the Kamchatka peninsula to the far east side of the Kuril Islands has been shown in shear wave waveform inversion [*Suzuki et al.,* 2016]. This feature is also observed across different updated models with patterns very similar to their study. Additionally, a low velocity anomaly is observed beneath south America in the lowermost mantle, with strong amplitude and consistent pattern across all 5 updated models. And at the same geographical locations, only SEWUCB_WM1 shows high amplitude low velocity anomaly, similar to the results in our updated models. Furthermore, significant heterogeneities have been identified in starting seismic tomography models [*Garnero et al.*, 2016] as well as forward modeling [*He and Wen*, 2012]. The results in our final updated models suggest very strong internal variations within LLSVP with similar patterns across different updated models. The strong velocity variation can not be explained by temperature effects alone and suggests compositional heterogeneity within LLSVP, which is consistent with the thermal and compositional heterogeneity that is predicted by geodynamic models [*Garnero and McNamara*, 2008]. The episodic entrainment of deeply subducted dense oceanic crust provides a plausible source for such heterogeneity [*Li et al*., 2014].

The consistency between the low velocity anomalies identified in forward and inversion studies [*Lekic et al*., 2012; *He and Wen*, 2012; *Suzuki et al*., 2016; *Nelson and Grand*, 2016] and our updated models provides confidence on the scale and amplitude in the updated models generated through the iterative forward update method. This method`s ability to resolve intermediate-scale features that are not consistent in starting models suggest higher resolving power than the starting tomography models.

To quantitatively measure the difference between different models, we show the standard deviation change of shear velocity perturbation between different models. We take the velocity perturbation from 5 different models and calculate the standard deviation for each of the cells and we plot the standard deviation map in a constant depth slice at 2800 km depth. The top panel shows the standard deviation map for all starting tomography models STD_{Before}, the middle panel shows the standard deviation map for all updated models STD_{After} and we further calculate the standard deviation model difference by subtracting $STD_{After} - STD_{Before}$, which is shown in the bottom panel. Figure 5.9c shows significant standard deviation reduction at locations that are associated with the Africa LLSVP and Pacific LLSVP, suggesting converged velocity between different starting models. Additionally, from Figure 5.9a to Figure 5.9b, the standard deviation has dropped significantly at global scale, and in Figure 5.9b, the standard deviations we see are almost uniformly distributed throughout the global, suggesting the updated models from iterative forward tomography are converging towards a similar solution.

During the update model, we document the residual travel time standard deviation reduction for each step in Table 5.3. The total residual travel time standard deviation reduction relative to the final updated model as well as PREM are also calculated in the table. From Table 5.3, by comparing the standard deviation reduction between step 1 and step 2, most of the standard deviation reductions are concentrated in step 1. The phases used in the step 1 update include S, SS, SSS, which are more than half of our dataset. The long travel path they go through in the upper mantle, especially for SS and SSS, carries significant travel time anomaly, which is accounted for in the step 1 update. The total

residual travel time standard deviation reduction is at the scale of 12% for most of the models relative to the starting tomography model and 66% relative to PREM. This suggests that most of the travel time anomaly is already accounted for by the starting tomography model we employ, and the iterative forward tomography process additionally accounted for 12% of the standard deviation. The small variance between the standard deviation reduction value also suggests the iterative forward update does not depend on the starting tomography model.

5.5 Conclusion

With the high-quality seismic dataset that we have created from the Empirical Wavelet construction method as well as Virtual Station stacking method, a large seismic dataset with more than 255K records with comprehensive measurements (travel time, waveform information et al.) has been documented. With this dataset, we apply layer stripping iterative forward tomography to iteratively update a starting model to account for the travel time differences between observed travel times in our dataset and predicted travel times from the updated model.

We introduce the layer stripping technique to update the update in a top-to-bottom manner to avoid structure contamination between the surface and the lowermost mantle. To accommodate this, our model is divided into two layers and a 2-step updating scheme is implemented where the top layer is updated first. In subsequent updates, the top layer is kept frozen and only the bottom layer is allowed to change. One of the challenges is that by freezing the top layers, the unexplained residual error from the top layer is attributed to the bottom layer, especially for cases where the top layer is not well accounted for. To address this issue, we used seismic tomography models as our starting

models. Global seismic tomography models often use phase velocity measurements from surface waves, which are more sensitive to shallow structures. With a starting seismic tomography model, a large portion of the residual standard deviation is accounted for and the accumulation effect is minimized.

During each step, the model is iteratively updated by taking the updated model from the previous iteration as the new starting model in current iteration. During each iteration, we implemented the RMS heterogeneity-based weights in assigning travel time residuals along the path of each individual record. After assigning all residual travel time, we convert the residual travel times within each cell into velocity perturbations and a Gaussian radius-based weighting scheme is applied to calculate the weighted average velocity perturbation. To improve the azimuthal coverage, especially in the lower mantle, we adapted the Gaussian radius based on the number of unique azimuthal samplings for each cell. However, with the Gaussian weight, a larger radius usually means stronger smoothing and smearing in the updated model. For locations where the azimuthal coverage is sparse, the tradeoff between better azimuthal coverage and stronger smoothing is not avoidable.

Five global seismic tomography models have been used as starting models in the iterative forward update method. By comparing the differences between the starting models and the updated models, we have identified consistent features in both pattern and amplitude in the updated models. This consistency suggests that the iterative forward updating scheme is updating different starting models into a convergent model. The consistency in residual travel time standard deviation reduction also suggests our method does not depend on starting model. We have identified uniform low velocity anomalies in

the updated models at several different locations in the lowermost mantle, which is consistent with results presented in forward and inverse studies on the same regions. Our final updated models also show strong internal variations within LLSVPs, especially the Pacific LLSVP. The strong velocity variation suggests compositional contributions, as predicted from geodynamics models. The final updated models` ability to resolve the intermediate-scale structures shows better resolving power than the starting tomography used.

5.6 References

Artemieva, I. M., and W. D. Mooney (2001), Thermal thickness and evolution of Precambrian lithosphere: A global study, *Journal of Geophysical Research: Solid Earth (1978–2012)*, *106*(B8), 16387–16414, doi:10.1029/2000JB900439.

Auer, L., L. Boschi, T. W. Becker, T. N. Meyer, and D. Giardini (2014), Savani: A variable resolution whole-mantle model of anisotropic shear velocity variations based on multiple data sets, *Journal of Geophysical Research: Solid Earth*, *119*(4), 3006–3034, doi:10.1002/2013JB010773.

Becker, T., and L. Boschi (2002), A comparison of tomographic and geodynamic mantle models, *Geochem. Geophys. Geosyst.*, 1–48.

Crotwell, H. P., T. J. Owens, and J. Ritsema (1999), The TauP Toolkit: Flexible Seismic Travel-time and Ray-path Utilities, *Seismological Research Letters*, *70*(2), 154–160, doi:10.1785/gssrl.70.2.154.

Durand, S., E. Debayle, Y. Ricard, C. Zaroli, and S. Lambotte (2017), Confirmation of a change in the global shear velocity pattern at around 1000 km depth, *Geophysical Journal International*, *211*(3), 1628–1639, doi:10.1093/gji/ggx405.

Dziewonski, A. M., and D. L. Anderson (1981), Preliminary reference Earth model, *Physics of the Earth and Planetary Interiors*, *25*(4), 297–356, doi:10.1016/0031-9201(81)90046-7.

French, S. W., and B. Romanowicz (2015), Broad plumes rooted at the base of the Earth's mantle beneath major hotspots, *Nature*, *525*(7567), 95–99, doi:10.1038/nature14876.

Grand, S. P. (2002), Mantle shear–wave tomography and the fate of subducted slabs, *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *360*(1800), 2475–2491, doi:10.1098/rsta.2002.1077.

Garnero, E. J., and A. K. McNamara (2008), Structure and Dynamics of Earth's Lower Mantle, *Science*, *320*(5876), 626–628, doi:10.1126/science.1148028.

Garnero, E. J., A. K. McNamara, and S.-H. Shim (2016), Continent-sized anomalous zones with low seismic velocity at the base of Earth's mantle, *Nature Geosci*, *9*(7), 481–489, doi:10.1038/ngeo2733.

He, Y., L. Wen, and T. Zheng (2006), Geographic boundary and shear wave velocity structure of the “Pacific anomaly” near the core–mantle boundary beneath western Pacific, *Earth and Planetary Science Letters*, *244*(1-2), 302–314, doi:10.1016/j.epsl.2006.02.007.

He, Y., and L. Wen (2012), Geographic boundary of the “Pacific Anomaly” and its geometry and transitional structure in the north, *Journal of Geophysical Research: Solid Earth (1978–2012)*, *117*(B9), doi:10.1029/2012JB009436.

Houser, C., G. Masters, P. Shearer, and G. Laske (2008), Shear and compressional velocity models of the mantle from cluster analysis of long-period waveforms, *Geophysical Journal International*, *174*(1), 195–212, doi:10.1111/j.1365-246X.2008.03763.x.

Kennett, B. L. N., and O. Gudmundsson (1996), Ellipticity corrections for seismic phases, *Geophysical Journal International*, *127*(1), 40–48, doi:10.1111/j.1365-246X.1996.tb01533.x.

Koelemeijer, P., J. Ritsema, A. Deuss, and H. J. Van Heijst (2016), SP12RTS: a degree-12 model of shear- and compressional-wave velocity for Earth's mantle, *Geophysical Journal International*, *204*(2), 1024–1039, doi:10.1093/gji/ggv481.

Kustowski, B., G. Ekström, and A. M. Dziewonski (2008), Anisotropic shear-wave velocity structure of the Earth's mantle: A global model, *Journal of Geophysical Research: Solid Earth (1978–2012)*, *113*(B6), B06306, doi:10.1029/2007JB005169.

Lai, H., E. J. Garnero, S. P. Grand, R. W. Porritt, and T. W. Becker (2019), Global Travel Time Data Set From Adaptive Empirical Wavelet Construction, *Geochem. Geophys. Geosyst.*, *20*(5), 2175–2198, doi:10.1029/2018GC007905.

Lekic, V., S. Cottaar, A. Dziewonski, and B. Romanowicz (2012), Cluster analysis of global lower mantle tomography: A new class of structure and implications for chemical heterogeneity, *Earth and Planetary Science Letters*, *357-358*, 68–77, doi:10.1016/j.epsl.2012.09.014.

Li, C., R. D. van der Hilst, E. R. Engdahl, and S. Burdick (2008), A new global model for P wave speed variations in Earth's mantle, *Geochem. Geophys. Geosyst.*, *9*(5), n/a–n/a, doi:10.1029/2007GC001806.

Li, M., A. K. McNamara, and E. J. Garnero (2014), Chemical complexity of hotspots caused by cycling oceanic crust through mantle reservoirs, *Nature Geosci*, *7*(5), 366–370, doi:10.1038/ngeo2120.

Mégnin, C., and B. Romanowicz (2000), The three-dimensional shear velocity structure of the mantle from the inversion of body, surface and higher-mode waveforms, *Geophysical Journal International*, *143*(3), 709–728, doi:10.1046/j.1365-246X.2000.00298.x.

Moulik, P., and G. Ekström (2014), An anisotropic shear velocity model of the Earth's mantle using normal modes, body waves, surface waves and long-period waveforms, *Geophysical Journal International*, *199*(3), 1713–1738, doi:10.1093/gji/ggu356.

Nelson, P. L., and S. P. Grand (2018), Lower-mantle plume beneath the Yellowstone hotspot revealed by core waves, *Nature Geoscience*, *11*(4), 280–284, doi:10.1038/s41561-018-0075-y.

Panning, M., and B. Romanowicz (2006), A three-dimensional radially anisotropic model of shear velocity in the whole mantle, *Geophysical Journal International*, *167*(1), 361–379, doi:10.1111/j.1365-246X.2006.03100.x.

Ritsema, J., A. Deuss, H. J. Van Heijst, and J. H. Woodhouse (2011), S40RTS: a degree-40 shear-velocity model for the mantle from new Rayleigh wave dispersion, teleseismic traveltime and normal-mode splitting function measurements, *Geophysical Journal International*, *184*(3), 1223–1236, doi:10.1111/j.1365-246X.2010.04884.x.

Romanowicz, B. (1991), Seismic Tomography of The Earth`s Mantle, *Annu. Rev. Earth Planet. Sci.*, 1–23.

Simmons, N. A., A. M. Forte, L. Boschi, and S. P. Grand (2010), GyPSuM: A joint tomographic model of mantle density and seismic wave speeds, *J. Geophys. Res.*, *115*(B12), B12310, doi:10.1029/2010JB007631.

Sun, D., D. Helmberger, S. Ni, and D. Bower (2009), Direct measures of lateral velocity variation in the deep Earth, *J. Geophys. Res.*, *114*(B5), B05303, doi:10.1029/2008JB005873.

Suzuki, Y., K. Kawai, R. J. Geller, A. F. E. Borgeaud, and K. Konishi (2016), Waveform inversion for 3-D S-velocity structure of D″ beneath the Northern Pacific: possible evidence for a remnant slab and a passive plume, *Earth, Planets and Space 2016 68:1*, *68*(1), 198, doi:10.1186/s40623-016-0576-0.

Wang, Y. (2004), Mapping the geometry and geographic distribution of a very low velocity province at the base of the Earth's mantle, *J. Geophys. Res.*, *109*(B10), B10305–18, doi:10.1029/2003JB002674.

Table 5.1. Number of travel time measurements used in this study

Phase Name	#Records
S	123,946
SS	53,505
SSS	11,927
Sdiff	28,499
ScS	23,758
ScSScS	10,303
Virtual S	561
Virtual SS	1652
Virtual SSS	1767
Virtual Sdiff	438
Virtual ScS	464
Virtual ScSScS	734
All	257,555

Table 5.2. Data record weight associated with different phase

Phase	Relative Weight
S	1.0
SS	0.8
SSS	0.7
ScSScS	0.65
ScS	0.7
Sdiff	1.0

Table 5.3. Residual travel time standard deviation reduction in Step 1 update, Step 2 update, 2-steps update combined and overall reduction relative to PREM

Model Name	Step 1 $\sigma_{residual}$ Reduction	Step 2 $\sigma_{residual}$ Reduction	Total $\sigma_{residual}$ Reduction Relative to starting model	Overall $\sigma_{residual}$ Reduction relative to PREM
SAW24B16	45.49%	11.97%	12.45%	66.95%
GyPSuM S	26.55%	6.15%	12.65%	67.03%
S40RTS	33.61%	7.68%	12.72%	67.05%
SEMUCB_WM1	41.60%	11.08%	10.53%	66.23%
SP12RTS S	35.83%	7.45%	12.34%	66.91%

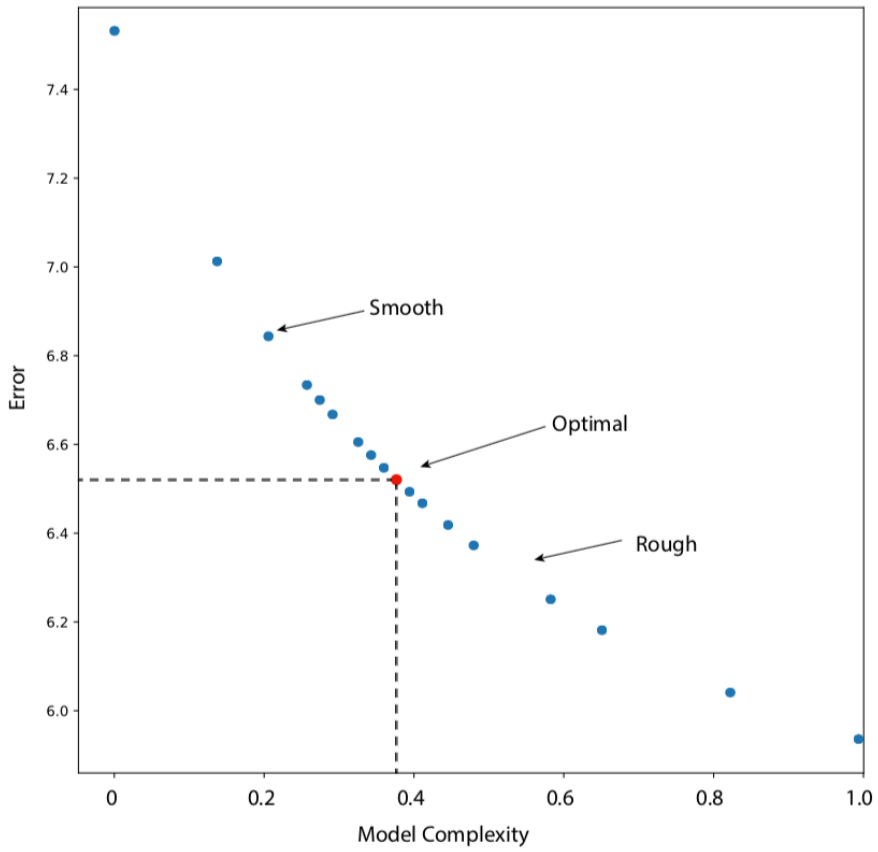

Figure 5.1. Tradeoff between model complexity (model parameter variance) and error (residual travel time variance)

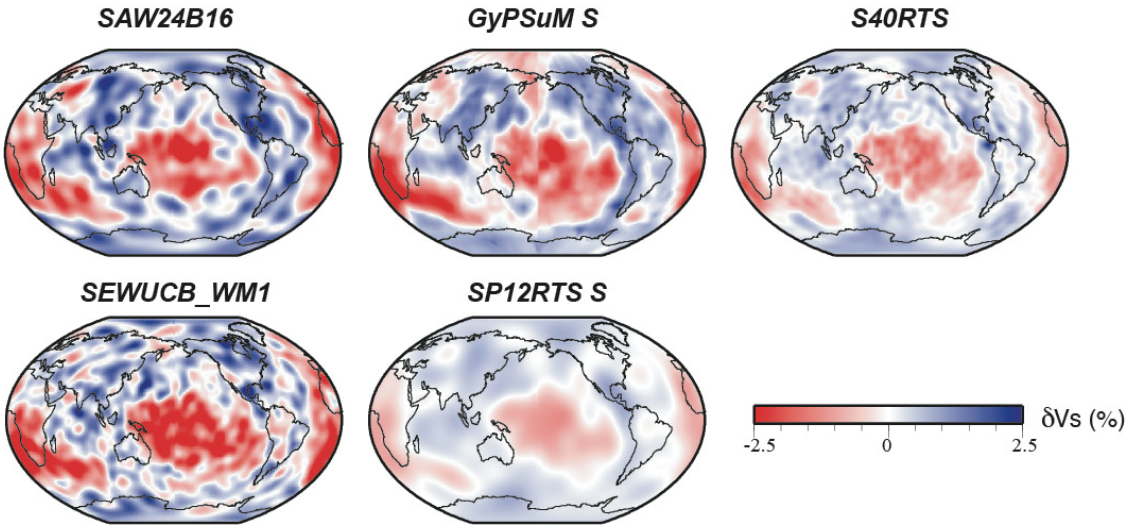

Figure 5.2. Horizontal slice of 5 global seismic tomography models plotted at 2800 km depth, including SAW24B16 [*Mégnin and Romanowicz*, 2000], GyPSuM-Shear wave model, [*Simmons et al.*, 2010], S40RTS [*Ritsema et al.*, 2011], SEWUCB_WM1[*French and Romanowicz*, 2015] and SP12RTS-Shear wave model [*Koelemeijer et al.*, 2016]

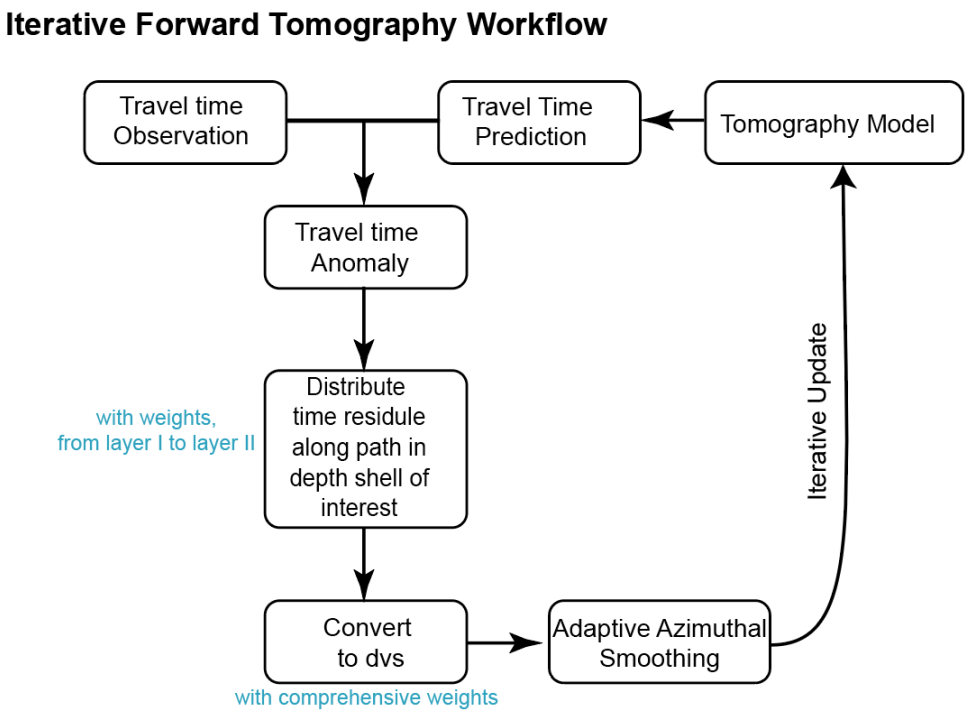

Figure 5.3. Iterative forward tomography update workflow

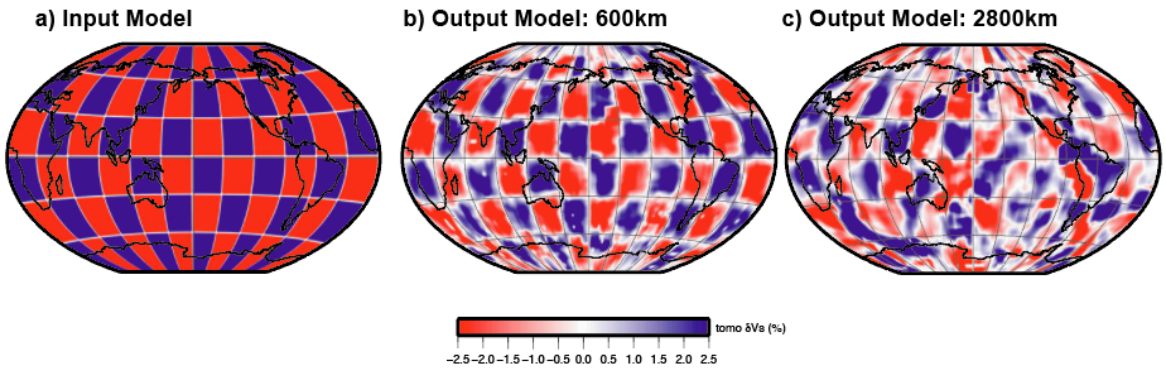

Figure 5.4. a) Checkerboard input model, with shear velocity amplitude alternating between 3% and -3% b) Output model at 600 km depth c) Output model at 2800 km depth

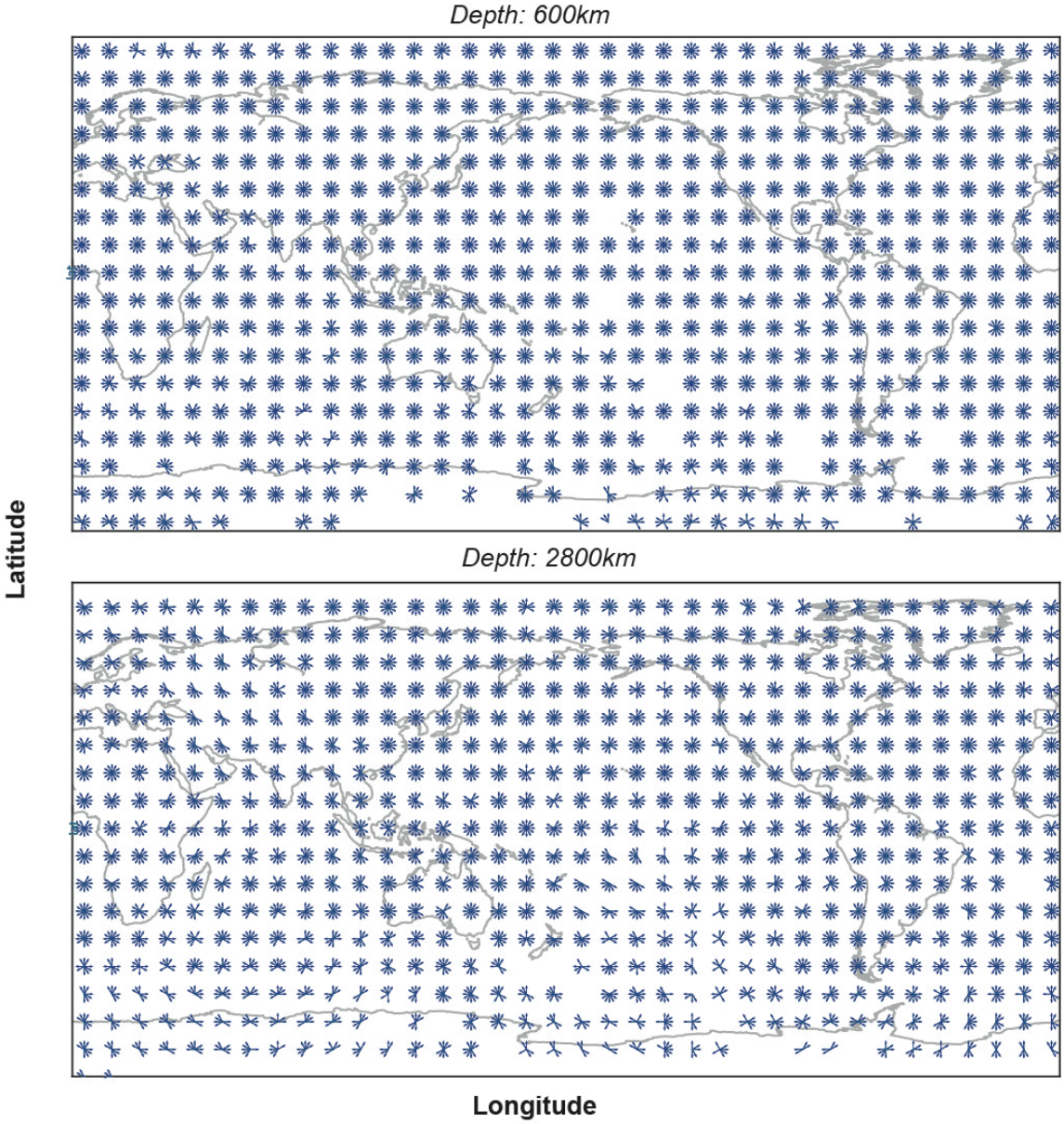

Figure 5.5. Azimuthal coverage plot of the dataset used in this study, plotted at 600 km (top panel) and 2800 km (bottom panel). The cells are down sampled to every 10 degree on both longitudinal and latitudinal direction for better visual effect. Each cell is represented by one arrow cluster, with each arrow corresponds to one azimuthal direction that has sampling.

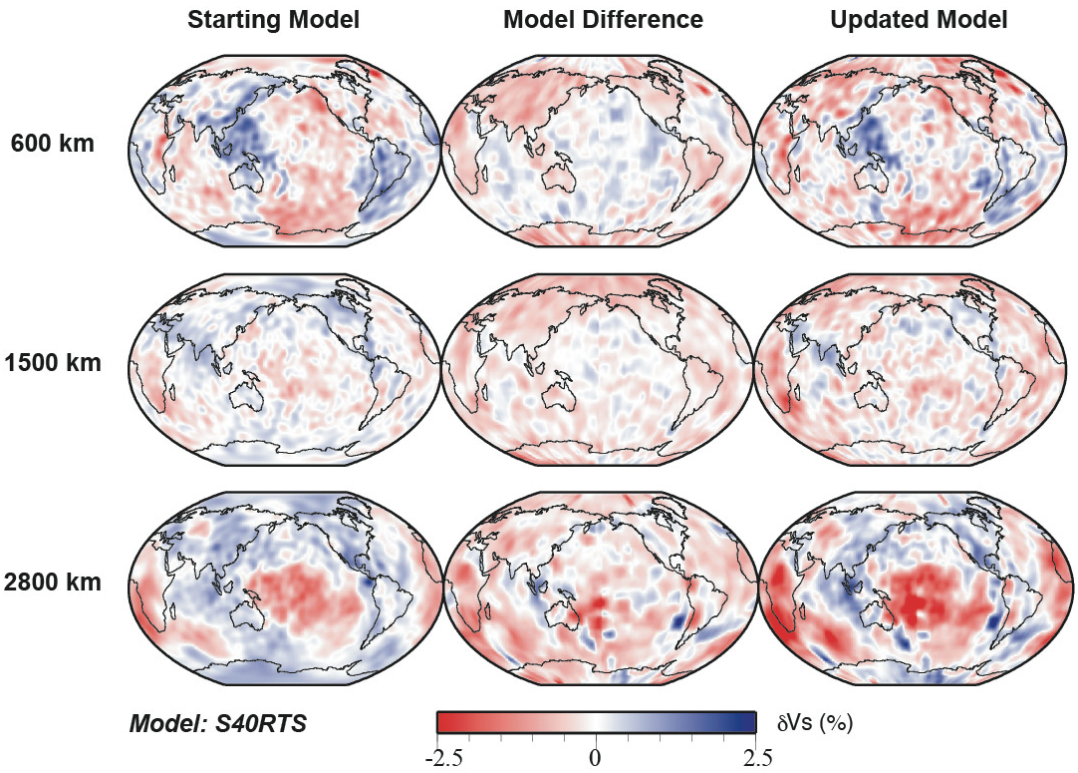

Figure 5.6. Horizontal slices of shear wave velocity perturbation for starting model (left panel), updated model (right panel) and their difference (middle panel) plotted at 600 km, 1500 km and 2800 km respectively. The shear velocity perturbation is relative to PREM and ranges from -2.5% to 2.5%, where blue means lower velocity and red means higher velocity

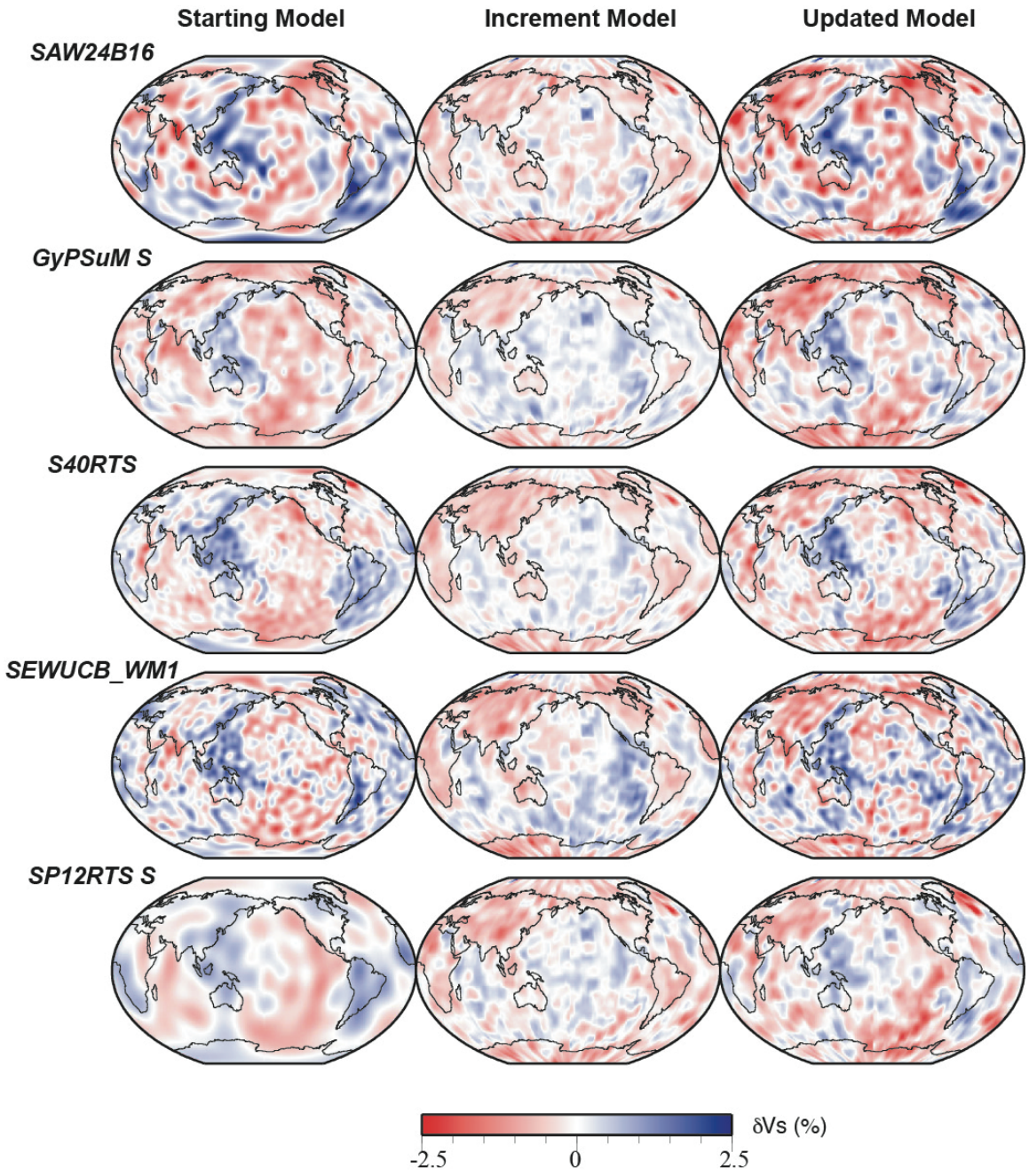

Figure 5.7. Final model update with 5 different starting seismic tomography models at 600 km depth. The left panel shows the starting tomography model, the right panel shows the updated final model and the middle panels shows the difference between the starting model and the updated model. The color represents the shear velocity perturbation relative to PREM in percentage from -2.5% to 2.5%.

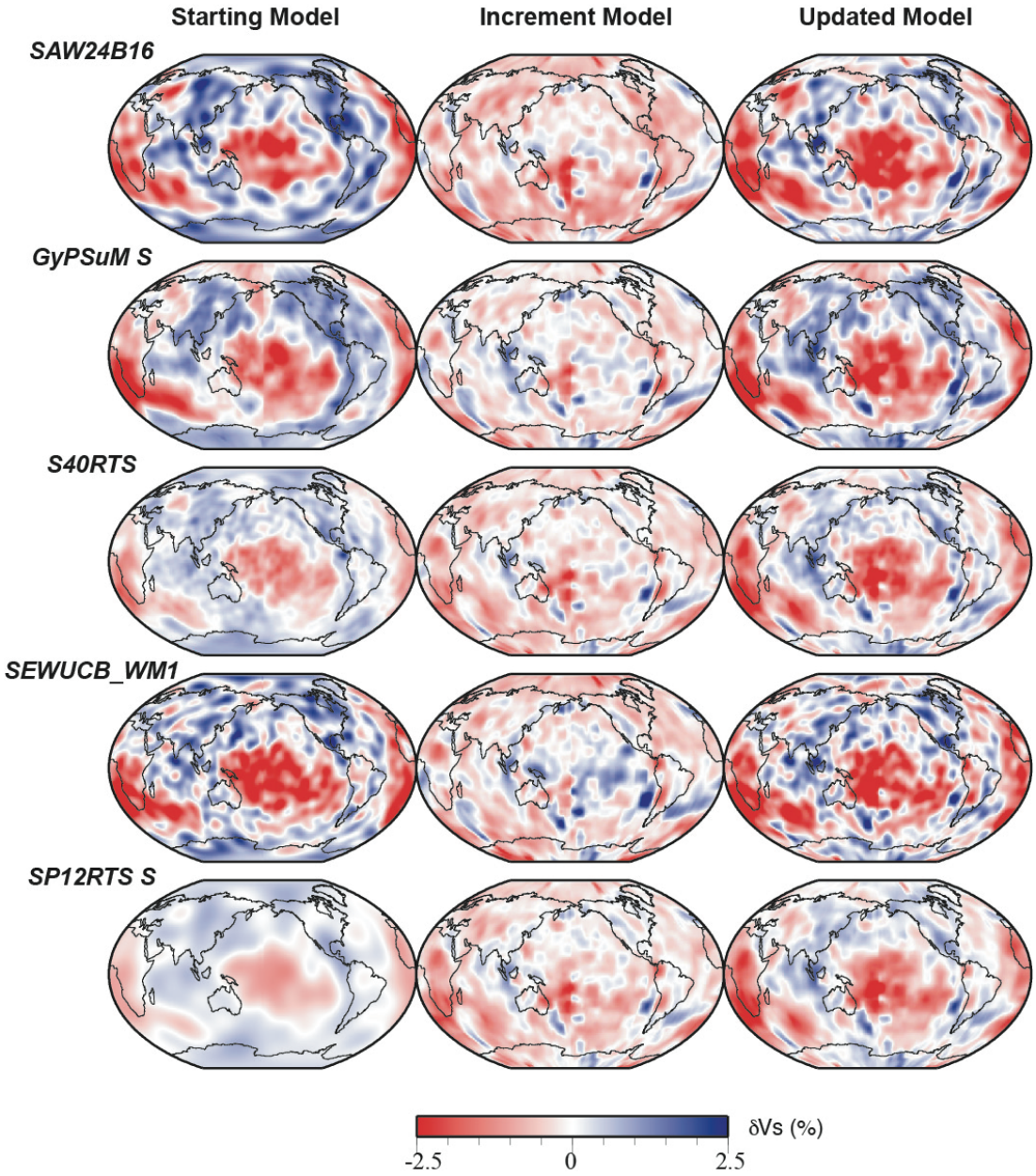

Figure 5.8. Final model update with 5 different starting seismic tomography models at 2800 km depth. The left panel shows the starting tomography model, the right panel shows the updated final model and the middle panels shows the difference between the starting model and the updated model. The color represents the shear velocity perturbation relative to PREM in percentage from -2.5% to 2.5%.

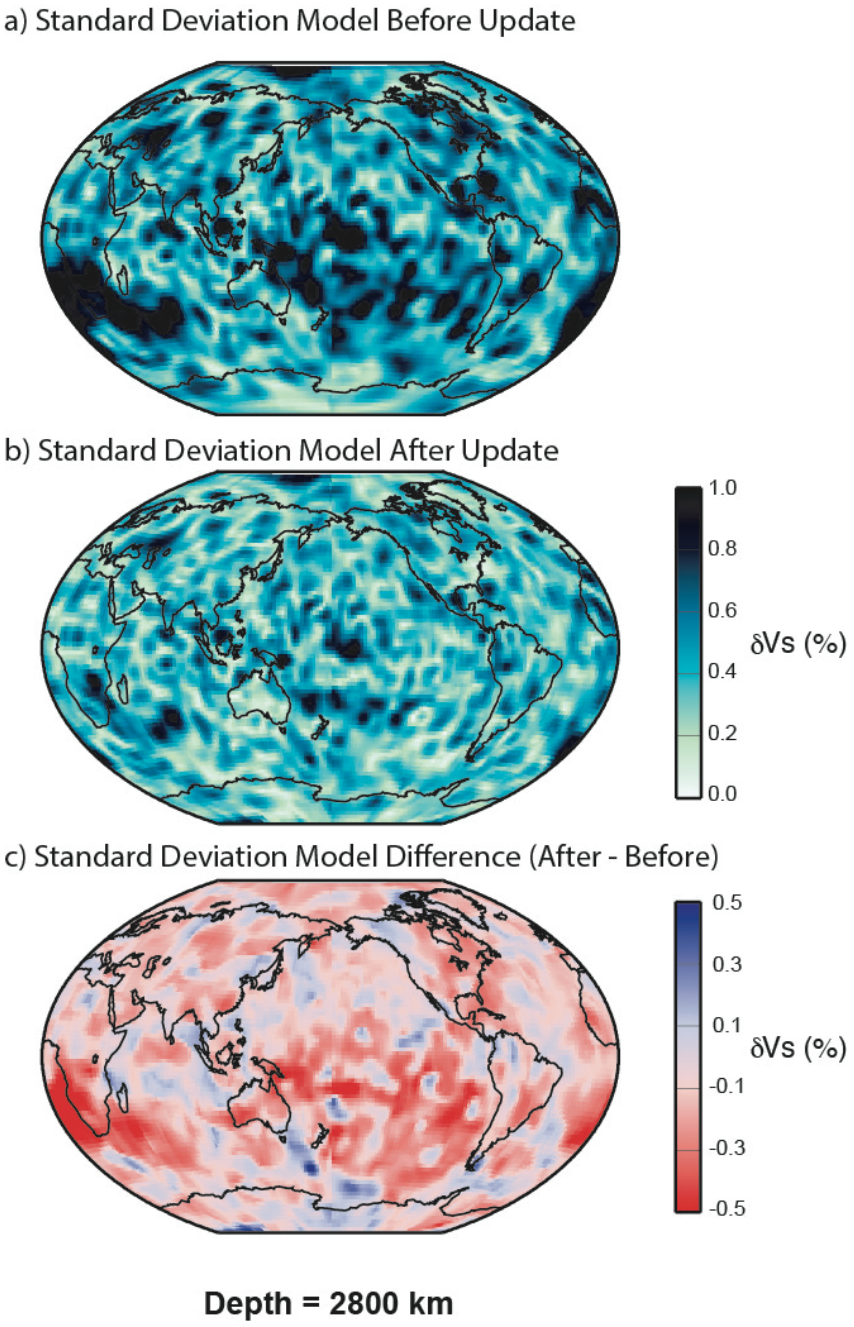

Figure 5.9. Standard deviation map showing the standard deviation of shear velocity of a) the 5 starting seismic tomography models, b) the 5 updated models from different starting models c) the standard deviation model difference by subtracting standard deviation model of updated model from standard deviation model of starting tomography models, where negative value suggesting standard deviation reduction

REFERENCES

Chapter 1.

Astiz, L. et al. (2014), The Array Network Facility Seismic Bulletin: Products and an Unbiased View of United States Seismicity, *Seismological Research Letters*, *85*(3), 576–593, doi:10.1785/0220130141.

Avants, M., T. Lay, and E. J. Garnero (2006), A new probe of ULVZ S-wave velocity structure: Array stacking of ScS waveforms, *Geophysical Research Letters*, *33*(7), doi:10.1029/2005GL024989.

Bréger, L., and B. Romanowicz (1998), Three-Dimensional Structure at the Base of the Mantle Beneath the Central Pacific, *Science*, *282*(5389), 718–720, doi:10.1126/science.282.5389.718.

Burke, K., B. Steinberger, T. H. Torsvik, and M. A. Smethurst (2008), Plume Generation Zones at the margins of Large Low Shear Velocity Provinces on the core–mantle boundary, *Earth and Planetary Science Letters*, *265*(1-2), 49–60, doi:10.1016/j.epsl.2007.09.042.

Durand, S., E. Debayle, Y. Ricard, C. Zaroli, and S. Lambotte (2017), Confirmation of a change in the global shear velocity pattern at around 1000 km depth, *Geophysical Journal International*, *211*(3), 1628–1639, doi:10.1093/gji/ggx405.

French, S. W., and B. Romanowicz (2015), Broad plumes rooted at the base of the Earth's mantle beneath major hotspots, *Nature*, *525*(7567), 95–99, doi:10.1038/nature14876.

Frost, D. A., S. Rost, N. D. Selby, and G. W. Stuart (2013), Detection of a tall ridge at the core–mantle boundary from scattered PKP energy, *Geophysical Journal International*, *195*(1), 558–574, doi:10.1093/gji/ggt242.

Garnero, E. J., and A. K. McNamara (2008), Structure and Dynamics of Earth's Lower Mantle, *Science*, *320*(5876), 626–628, doi:10.1126/science.1148028.

Garnero, E. J., A. K. McNamara, and S.-H. Shim (2016), Continent-sized anomalous zones with low seismic velocity at the base of Earth's mantle, *Nature Geosci*, *9*(7), 481–489, doi:10.1038/ngeo2733.

Garnero, E. J., and D. V. Helmberger (1996), Seismic detection of a thin laterally varying boundary layer at the base of the mantle beneath the central-Pacific, *Geophysical Research Letters*, *23*(9), 977–980, doi:10.1029/95GL03603.

Grand, S. P. (2002), Mantle shear–wave tomography and the fate of subducted slabs, *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *360*(1800), 2475–2491, doi:10.1098/rsta.2002.1077.

He, Y., and L. Wen (2009), Structural features and shear-velocity structure of the "Pacific Anomaly," *Journal of Geophysical Research: Solid Earth (1978–2012)*, *114*(B2), doi:10.1029/2008JB005814.

He, Y., and L. Wen (2012), Geographic boundary of the "Pacific Anomaly" and its geometry and transitional structure in the north, *Journal of Geophysical Research: Solid Earth (1978–2012)*, *117*(B9), doi:10.1029/2012JB009436.

Hutko, A. R., T. Lay, and J. Revenaugh (2009), Localized double-array stacking analysis of PcP: D″ and ULVZ structure beneath the Cocos plate, Mexico, central Pacific, and north Pacific, *Physics of the Earth and Planetary Interiors*, *173*(1-2), 60–74, doi:10.1016/j.pepi.2008.11.003.

Koelemeijer, P., J. Ritsema, A. Deuss, and H. J. Van Heijst (2016), SP12RTS: a degree-12 model of shear- and compressional-wave velocity for Earth's mantle, *Geophysical Journal International*, *204*(2), 1024–1039, doi:10.1093/gji/ggv481.

Kohler, M. D., J. E. Vidale, and P. M. Davis (1997), Complex scattering within D″ observed on the very dense Los Angeles Region Seismic Experiment Passive Array, *Geophysical Research Letters*, *24*(15), 1855–1858, doi:10.1029/97GL01823.

Lay, T., J. Hernlund, E. J. Garnero, and M. S. Thorne (2006), A Post-Perovskite Lens and D"" Heat Flux Beneath the Central Pacific, *Science*, *314*(5803), 1272–1276, doi:10.1126/science.1133280.

Lekic, V., S. Cottaar, A. Dziewonski, and B. Romanowicz (2012), Cluster analysis of global lower mantle tomography: A new class of structure and implications for chemical heterogeneity, *Earth and Planetary Science Letters*, *357-358*, 68–77, doi:10.1016/j.epsl.2012.09.014.

Masters, G., and G. Laske (2000), The Relative Behavior of Shear Velocity, Bulk Sound Speed, and Compressional Velocity in the Mantle: Implications for Chemical and Thermal Structure,, 1–28.

McNamara, A. K., E. J. Garnero, and S. Rost (2010a), Tracking deep mantle reservoirs with ultra-low velocity zones, *Earth and Planetary Science Letters*, *299*(1-2), 1–9, doi:10.1016/j.epsl.2010.07.042.

Mégnin, C., and B. Romanowicz (2000), The three-dimensional shear velocity structure of the mantle from the inversion of body, surface and higher-mode waveforms, *Geophysical Journal International*, *143*(3), 709–728, doi:10.1046/j.1365-246X.2000.00298.x.

Moore, M. M., E. J. Garnero, T. Lay, and Q. Williams (2004), Shear wave splitting and waveform complexity for lowermost mantle structures with low-velocity lamellae and transverse isotropy, *J. Geophys. Res.*, *109*(B2), 173–26, doi:10.1029/2003JB002546.

Mori, J., and D. V. Helmberger (1995), Localized boundary layer below the mid-Pacific velocity anomaly identified from a PcP precursor, *Journal of Geophysical Research: Solid Earth (1978–2012)*, *100*(B10), 20359–20365, doi:10.1029/95JB02243.

Ni, S. (2002), Sharp Sides to the African Superplume, *Science*, *296*(5574), 1850–1852, doi:10.1126/science.1070698.

Revenaugh, J., and R. Meyer (1997), Seismic Evidence of Partial Melt Within a Possibly Ubiquitous Low-Velocity Layer at the Base of the Mantle, *Science*, *277*(5326), 670–673, doi:10.1126/science.277.5326.670.

Ritsema, J., A. Deuss, H. J. Van Heijst, and J. H. Woodhouse (2011), S40RTS: a degree-40 shear-velocity model for the mantle from new Rayleigh wave dispersion, teleseismic traveltime and normal-mode splitting function measurements, *Geophysical Journal International*, *184*(3), 1223–1236, doi:10.1111/j.1365-246X.2010.04884.x.

Ritsema, J., T. Lay, E. J. Garnero, and H. Benz (1998), Seismic anisotropy in the lowermost mantle beneath the Pacific, *Geophysical Research Letters*.

Rost, S. (2002), Array seismology: Methods and applications, *Rev. Geophys.*, *40*(3), 1008, doi:10.1029/2000RG000100.

Russell, S. A., T. Lay, and E. J. Garnero (1998), Seismic evidence for small-scale dynamics in the lowermost mantle at the root of the Hawaiian hotspot,, *396*(6708), 255–258, doi:10.1038/24364.

Russell, S. A., T. Lay, and E. J. Garnero (1999), Small-scale lateral shear velocity and anisotropy heterogeneity near the core-mantle boundary beneath the central Pacific imaged using broadband ScS waves, *Journal of Geophysical Research: Solid Earth (1978–2012)*, *104*(B6), 13183–13199, doi:10.1029/1999JB900114.

Russell, S. J., and Norvig, P. (2016). Artificial intelligence: a modern approach. Malaysia; *Pearson Education Limited.*

Simmons, N. A., A. M. Forte, L. Boschi, and S. P. Grand (2010), GyPSuM: A joint tomographic model of mantle density and seismic wave speeds, *J. Geophys. Res.*, *115*(B12), B12310, doi:10.1029/2010JB007631.

Sun, D., D. Helmberger, S. Ni, and D. Bower (2009), Direct measures of lateral velocity variation in the deep Earth, *J. Geophys. Res.*, *114*(B5), B05303, doi:10.1029/2008JB005873.

Sun, D., E. Tan, D. Helmberger, and M. Gurnis (2007), Seismological support for the metastable superplume model, sharp features, and phase changes within the lower mantle, *PNAS*, *104*(22), 9151–9155, doi:10.1073/pnas.0608160104.

Suzuki, Y., K. Kawai, R. J. Geller, A. F. E. Borgeaud, and K. Konishi (2016), Waveform inversion for 3-D S-velocity structure of D″ beneath the Northern Pacific: possible evidence for a remnant slab and a passive plume, *Earth, Planets and Space 2016 68:1*, *68*(1), 198, doi:10.1186/s40623-016-0576-0.

Thorne, M. S., and E. J. Garnero (2004), Inferences on ultralow-velocity zone structure from a global analysis of SPdKSwaves, *J. Geophys. Res.*, *109*(B8), 421–22, doi:10.1029/2004JB003010.

Torsvik, T. H., K. Burke, B. Steinberger, S. J. Webb, and L. D. Ashwal (2010), Diamonds sampled by plumes from the core-mantle boundary, *Nature*, *466*(7304), 352–355, doi:10.1038/nature09216.

Zhao, C., E. J. Garnero, A. K. McNamara, N. Schmerr, and R. W. Carlson (2015), Seismic evidence for a chemically distinct thermochemical reservoir in Earth's deep mantle beneath Hawaii, *Earth and Planetary Science Letters*, *426*(C), 143–153, doi:10.1016/j.epsl.2015.06.012.

Chapter 2.

Astiz, L., Eakins, J. A., Martynov, V. G., Cox, T. A., Tytell, J., Reyes, J. C., Newman, R. L., Karasu, G. H., Mulder, T., White, M., Davis, G. A., Busby, R. W., Hafner, K., Meyer, J. C., and Vernon, F. L. (2014), The Array Network Facility Seismic Bulletin: Products and an Unbiased View of United States Seismicity, Seismo. Res. Lett., 85 (3), doi:10.1785/0220130141.

Bolton, H., and G. Masters (2001), Travel times of P and S from the global digital seismic networks: Implications for the relative variation of P and S velocity in the mantle, *J. Geophys. Res., 106*(B7), 13527–13540, doi:10.1029/2000JB900378.

Butler, R. (1979), Shear-wave travel times from SS, *Bulletin of the Seismological Society of America*, *69*(6), 1715–1732, doi:10.1121/1.1906594.

Choy, G. L., and P. G. Richards (1975), Pulse distortion and Hilbert transformation in multiply reflected and refracted body waves, *Bulletin of the Seismological Society of America*, *65*(1), 55–70, doi:10.1121/1.1906594.

Conder, J. A. (2015), Fitting Multiple Bell Curves Stably and Accurately to a Time Series as Applied to Hubbert Cycles or Other Phenomena, *Mathematical Geosciences*, *47*(6), 663–678, doi:10.1007/s11004-014-9557-7.

Cottaar, S., and V. Lekic (2016), Morphology of seismically slow lower-mantle structures, *Geophysical Journal International*, *207*(2), 1122–1136, doi:10.1093/gji/ggw324.

Crotwell, H. P., T. J. Owens, and J. Ritsema (1999), The TauP Toolkit: Flexible seismic travel-time and ray-path utilities, *Seismological Research Letters 70*, 154–160.

Davies, C., M. Pozzo, D. Gubbins, and D. Alfè (2015a), Constraints from material properties on the dynamics and evolution of Earth's core, *Nature Geosci*, *8*(9), 678–685, doi:10.1038/ngeo2492.

Davies, D. R., S. Goes, and H. C. P. Lau (2015b), Thermally Dominated Deep Mantle LLSVPs: A Review, in *The Earth's Heterogeneous Mantle*, pp. 441–477, Springer International Publishing, Cham.

Dziewonski, A. M., and D. L. Anderson (1981), Preliminary reference Earth model, Phys. Earth Planet. Inter., 25, 297–356, doi:10.1016/0031-9201(81)90046-7.

Dziewonski, A. M., T. A. Chou, and J. H. Woodhouse (1981), Determination of earthquake source parameters from waveform data for studies of global and regional seismicity, *Journal of Geophysical Research: Solid Earth (1978–2012)*, *86*(B4), 2825–2852, doi:10.1029/JB086iB04p02825.

Ekström, G., M. Nettles, and A. M. Dziewonski (2012), The global CMT project 2004–2010: Centroid-moment tensors for 13,017 earthquakes, *Physics of the Earth and Planetary Interiors*, *200-201*, 1–9, doi:10.1016/j.pepi.2012.04.002.

Futterman, W. I. (1962), Dispersive body waves, *Journal of Geophysical Research: Solid Earth (1978–2012)*, *67*(13), 5279–5291, doi:10.1029/JZ067i013p05279.

Garnero, E. J., A. K. McNamara, and S.-H. Shim (2016), Continent-sized anomalous zones with low seismic velocity at the base of Earth's mantle, *Nature Geosci*, *9*(7), 481–489, doi:10.1038/ngeo2733.

Garnero, E. J., and A. K. McNamara (2008), Structure and Dynamics of Earth's Lower Mantle, *Science*, *320*(5876), 626–628, doi:10.1126/science.1148028.

Golos, M., Fang, H., Yao, H., Zhang, H., Burdick, S., Vernon, F., Schaeffer, A., Lebedev, S., and van der Hilst, R. D. (2018), Shear Wave Tomography Beneath the United States Using a Joint Inversion of Surface and Body Waves, J. Geophys. Res., Solid Earth, doi:10.1029/2017JB014894.

Gu, Y. H., A. L. Lerner-Lam, A. M. Dziewonski, and G. Ekström (2005), Deep structure and seismic anisotropy beneath the East Pacific Rise, Earth Planet. Sci. Lett., 232, 259–272, doi:10.1016/j.epsl.2005.01.019.

Houser, C., G. Masters, P. Shearer, and G. Laske (2008), Shear and compressional velocity models of the mantle from cluster analysis of long-period waveforms, *Geophysical Journal International*, *174*(1), 195–212, doi:10.1111/j.1365-246X.2008.03763.x.

Inoue, H., Y. Fukao, K. Tanabe, and Y. Ogata (1990), Whole mantle P-wave travel time tomography, *Physics of the Earth and Planetary Interiors*, *59*(4), 294–328, doi:10.1016/0031-9201(90)90236-Q.

Kennett, B. L. N., E. R. Engdahl, and R. Buland (1995), Constraints on seismic velocities in the Earth from traveltimes, *Geophysical Journal International*, *122*(1), 108–124, doi:10.1111/j.1365-246X.1995.tb03540.x.

Kustowski, B., G. Ekström, and A. M. Dziewonski (2008), Anisotropic shear-wave velocity structure of the Earth's mantle: A global model, *Journal of Geophysical Research: Solid Earth (1978–2012)*, *113*(B6), B06306, doi:10.1029/2007JB005169.

Laske, G., G. Masters, Z. Ma, and M. Pasyanos (2013), Update on CRUST1.0—A 1-degree Global Model of Earth's Crust, Geophys. Res. Abstr.,15, EGU2013-2658.

Lekic, V., S. Cottaar, A. Dziewonski, and B. Romanowicz (2012), Cluster analysis of global lower mantle tomography: A new class of structure and implications for chemical heterogeneity, *Earth and Planetary Science Letters*, *357-358*, 68–77, doi:10.1016/j.epsl.2012.09.014.

Lou, X., and S. van der Lee (2014), Observed and predicted North American teleseismic delay times, *Earth and Planetary Science Letters*, *402*, 6–15, doi:10.1016/j.epsl.2013.11.056.

Lou, X., Van Der Lee, S., and Lloyd, S. (2013), AIMBAT: A python/matplotlib tool for measuring teleseismic arrival times, *Seismological Research Letters, 84*(1), 85-93.

Ni, S., and Helmberger, D. V. (2003), Ridge-like lower mantle structure beneath South Africa, *J. Geophys. Res., 108*, 2094, doi:10.1029/2001JB001545.

Pavlis, G. L. and Vernon, F. L. (2010), Array processing of teleseismic body waves with the USArray, *Computers and Geosciences, 36*(7), 910–920.

Ritsema, J., A. Deuss, H. J. Van Heijst, and J. H. Woodhouse (2011), S40RTS: a degree-40 shear-velocity model for the mantle from new Rayleigh wave dispersion, teleseismic traveltime and normal-mode splitting function measurements, *Geophysical Journal International*, *184*(3), 1223–1236, doi:10.1111/j.1365-246X.2010.04884.x.

Ritsema, J., and H. J. van Heijst (2002), Constraints on the correlation of P- and S-wave velocity heterogeneity in the mantle from P, PP, PPP and PKPab traveltimes, *Geophysical Journal International*, *149*(2), 482–489, doi:10.1046/j.1365-246X.2002.01631.x.

Rost, S. (2002), Array seismology: Methods and applications, *Rev. Geophys.*, *40*(3), 1008, doi:10.1029/2000RG000100.

Russell, S. J., and Norvig, P. (2016). Artificial intelligence: a modern approach. Malaysia; *Pearson Education Limited.*

Schaff, D. P., and F. Waldhauser (2005), Waveform Cross-Correlation-Based Differential Travel-Time Measurements at the Northern California Seismic Network, *Bulletin of the Seismological Society of America*, *95*(6), 2446–2461, doi:10.1785/0120040221.

Simmons, N. A., A. M. Forte, L. Boschi, and S. P. Grand (2010), GyPSuM: A joint tomographic model of mantle density and seismic wave speeds, *J. Geophys. Res.*, *115*(B12), B12310, doi:10.1029/2010JB007631.

To, A., Romanowicz, B., Capdeville, Y., and Takeuchi, N. (2005), 3D effects of sharp boundaries at the borders of the African and Pacific Superplumes: Observation and modeling, *Earth Planet. Sci. Lett., 233*, 137-153, doi:10.1016/j.epsl.2005.01.037.

van der Hilst, R. D., S. Widiyantoro, and E. R. Engdahl (1997), Evidence for deep mantle circulation from global tomography, *Nature*, *386*(6625), 578–584, doi:10.1038/386578a0.

Vandecar, J. C., and R. S. Crosson (1990), Determination of teleseismic relative phase arrival times using multi-channel cross-correlation and least squares, *Bulletin of the Seismological Society of America*, *80*(1), 150–169, doi:10.1029/JB091iB14p13873.

Woodhouse, J. H., and A. M. Dziewonski (1984), Mapping the upper mantle: Three-dimensional modeling of earth structure by inversion of seismic waveforms, *Journal of Geophysical Research: Solid Earth (1978–2012)*, *89*(B7), 5953–5986.

Woodward, R. L., and G. Masters (1991), Global upper mantle structure from long-period differential travel times, J. Geophys. Res., 96(B4), 6351–6377, doi:10.1029/90JB01445.

Chapter 3.

Ahmed, N. K., Atiya, A. F., Gayar, N. E., & El-Shishiny, H. (2010). An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, *29*(5-6), 594-621.

Astiz, L. et al. (2014), The Array Network Facility Seismic Bulletin: Products and an Unbiased View of United States Seismicity, *Seismological Research Letters*, *85*(3), 576–593, doi:10.1785/0220130141.

Bontempi, G., Taieb, S. B., & Le Borgne, Y. A. (2012, July). Machine learning strategies for time series forecasting. In *European business intelligence summer school* (pp. 62-77). Springer, Berlin, Heidelberg.

Breiman, L. (1996), Bagging predictors, *Mach Learning*, *24*(2), 123–140, doi:10.1007/BF00058655.

Freund, Y., and R. E. Schapire (1997), A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting, *Journal of Computer and System Sciences*, *55*(1), 119–139, doi:10.1006/jcss.1997.1504.

Ke, G., Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu (2017), LightGBM: A Highly Efficient Gradient Boosting Decision Tree, *papers.nips.cc*, 3146–3154.

Lai, H., E. J. Garnero, S. P. Grand, R. W. Porritt, and T. W. Becker (2019), Global Travel Time Data Set From Adaptive Empirical Wavelet Construction, *Geochem. Geophys. Geosyst.*, *20*(5), 2175–2198, doi:10.1029/2018GC007905.

Le, Q. V., Ngiam, J., Coates, A., Lahiri, A., Prochnow, B., & Ng, A. Y. (2011, June). On optimization methods for deep learning. In *Proceedings of the 28th International Conference on International Conference on Machine Learning* (pp. 265-272). Omnipress.

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. MIT press.

Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., & McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations* (pp. 55-60).

Meng, Q., G. Ke, T. Wang, W. Chen, Q. Ye, Z.-M. Ma, and T.-Y. Liu (2016), A Communication-Efficient Parallel Algorithm for Decision Tree, *papers.nips.cc*, 1279–1287.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *Journal of machine learning research*, *12*(Oct), 2825-2830.

Rosten, E., & Drummond, T. (2006, May). Machine learning for high-speed corner detection. In *European conference on computer vision* (pp. 430-443). Springer, Berlin, Heidelberg.

Scholkopf, B., & Smola, A. J. (2001). *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press.

Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms. In *Advances in neural information processing systems* (pp. 2951-2959).

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2818-2826).

Vedaldi, A., & Fulkerson, B. (2010, October). VLFeat: An open and portable library of computer vision algorithms. In *Proceedings of the 18th ACM international conference on Multimedia* (pp. 1469-1472). ACM.

Zhang, H., Si, S., & Hsieh, C. J. (2017). GPU-acceleration for Large-scale Tree Boosting. *arXiv preprint arXiv:1706.08359*.

Chapter 4

Artemieva, I. M., and W. D. Mooney (2001), Thermal thickness and evolution of Precambrian lithosphere: A global study, *Journal of Geophysical Research: Solid Earth (1978–2012)*, *106*(B8), 16387–16414, doi:10.1029/2000JB900439.

Auer, L., L. Boschi, T. W. Becker, T. N. Meyer, and D. Giardini (2014), Savani: A variable resolution whole-mantle model of anisotropic shear velocity variations based on multiple data sets, *Journal of Geophysical Research: Solid Earth*, *119*(4), 3006–3034, doi:10.1002/2013JB010773.

Becker, T., and L. Boschi (2002), A comparison of tomographic and geodynamic mantle models, *Geochem. Geophys. Geosyst.*, 1–48.

Bréger, L., and B. Romanowicz (1998), Three-Dimensional Structure at the Base of the Mantle Beneath the Central Pacific, *Science*, *282*(5389), 718–720, doi:10.1126/science.282.5389.718.

Cottaar, S., and V. Lekic (2016), Morphology of seismically slow lower-mantle structures, *Geophysical Journal International*, *207*(2), 1122–1136, doi:10.1093/gji/ggw324.

Durand, S., E. Debayle, Y. Ricard, C. Zaroli, and S. Lambotte (2017), Confirmation of a change in the global shear velocity pattern at around 1000 km depth, *Geophysical Journal International*, *211*(3), 1628–1639, doi:10.1093/gji/ggx405.

Dziewonski, A. M., and D. L. Anderson (1981), Preliminary reference Earth model, *Physics of the Earth and Planetary Interiors*, *25*(4), 297–356, doi:10.1016/0031-9201(81)90046-7.

Dziewonski, A. M., T. A. Chou, and J. H. Woodhouse (1981), Determination of earthquake source parameters from waveform data for studies of global and regional seismicity, *Journal of Geophysical Research: Solid Earth (1978–2012)*, *86*(B4), 2825–2852, doi:10.1029/JB086iB04p02825.

Ekström, G., M. Nettles, and A. M. Dziewonski (2012), The global CMT project 2004–2010: Centroid-moment tensors for 13,017 earthquakes, *Physics of the Earth and Planetary Interiors*, *200-201*, 1–9, doi:10.1016/j.pepi.2012.04.002.

Flanagan, M. P., and P. M. Shearer (1998), Global mapping of topography on transition zone velocity discontinuities by stacking SS precursors, *Journal of Geophysical Research: Solid Earth (1978–2012)*, *103*(B2), 2673–2692, doi:10.1029/97JB03212.

French, S. W., and B. Romanowicz (2015), Broad plumes rooted at the base of the Earth's mantle beneath major hotspots, *Nature*, *525*(7567), 95–99, doi:10.1038/nature14876.

Frost, D. A., and S. Rost (2014), The P-wave boundary of the Large-Low Shear Velocity Province beneath the Pacific, *Earth and Planetary Science Letters*, *403*, 380–392, doi:10.1016/j.epsl.2014.06.046.

Frost, D. A., S. Rost, N. D. Selby, and G. W. Stuart (2013), Detection of a tall ridge at the core–mantle boundary from scattered PKP energy, *Geophysical Journal International*, *195*(1), 558–574, doi:10.1093/gji/ggt242.

Futterman, W. I. (1962), Dispersive body waves, *Journal of Geophysical Research: Solid Earth (1978–2012)*, *67*(13), 5279–5291, doi:10.1029/JZ067i013p05279.

Garnero, E. J., A. K. McNamara, and S.-H. Shim (2016), Continent-sized anomalous zones with low seismic velocity at the base of Earth's mantle, *Nature Geosci*, *9*(7), 481–489, doi:10.1038/ngeo2733.

Grand, S. P. (2002), Mantle shear–wave tomography and the fate of subducted slabs, *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *360*(1800), 2475–2491, doi:10.1098/rsta.2002.1077.

Grand, S. P., and D. V. Helmberger (1984), Upper mantle shear structure of North America, *Geophysical Journal International*, *76*(2), 399–438, doi:10.1111/j.1365-246X.1984.tb05053.x.

Houser, C., G. Masters, P. Shearer, and G. Laske (2008), Shear and compressional velocity models of the mantle from cluster analysis of long-period waveforms, *Geophysical Journal International*, *174*(1), 195–212, doi:10.1111/j.1365-246X.2008.03763.x.

Kennett, B. L. N., and O. Gudmundsson (1996), Ellipticity corrections for seismic phases, *Geophysical Journal International*, *127*(1), 40–48, doi:10.1111/j.1365-246X.1996.tb01533.x.

Koelemeijer, P., J. Ritsema, A. Deuss, and H. J. Van Heijst (2016), SP12RTS: a degree-12 model of shear- and compressional-wave velocity for Earth's mantle, *Geophysical Journal International*, *204*(2), 1024–1039, doi:10.1093/gji/ggv481.

Kustowski, B., G. Ekström, and A. M. Dziewonski (2008), Anisotropic shear-wave velocity structure of the Earth's mantle: A global model, *Journal of Geophysical Research: Solid Earth (1978–2012)*, *113*(B6), B06306, doi:10.1029/2007JB005169.

Lekic, V., S. Cottaar, A. Dziewonski, and B. Romanowicz (2012), Cluster analysis of global lower mantle tomography: A new class of structure and implications for chemical heterogeneity, *Earth and Planetary Science Letters*, *357-358*, 68–77, doi:10.1016/j.epsl.2012.09.014.

Lai, H., E. J. Garnero, S. P. Grand, R. W. Porritt, and T. W. Becker (2019), Global Travel Time Data Set From Adaptive Empirical Wavelet Construction, *Geochem. Geophys. Geosyst.*, *20*(5), 2175–2198, doi:10.1029/2018GC007905.

Li, C., R. D. van der Hilst, E. R. Engdahl, and S. Burdick (2008), A new global model for P wave speed variations in Earth's mantle, *Geochem. Geophys. Geosyst.*, *9*(5), n/a–n/a, doi:10.1029/2007GC001806.

Moulik, P., and G. Ekström (2014), An anisotropic shear velocity model of the Earth's mantle using normal modes, body waves, surface waves and long-period waveforms, *Geophysical Journal International*, *199*(3), 1713–1738, doi:10.1093/gji/ggu356.

Ni, S. (2002), Sharp Sides to the African Superplume, *Science*, *296*(5574), 1850–1852, doi:10.1126/science.1070698.

Ni, S., and D. V. Helmberger (2003a), Further constraints on the African superplume structure, *Physics of the Earth and Planetary Interiors*, *140*(1-3), 243–251, doi:10.1016/j.pepi.2003.07.011.

Ni, S., and D. V. Helmberger (2003b), Ridge-like lower mantle structure beneath South Africa, *J. Geophys. Res.*, *108*(B2), 3433–14, doi:10.1029/2001JB001545.

Ni, S., D. V. Helmberger, and J. Tromp (2005), Three-dimensional structure of the African superplume from waveform modelling, *Geophysical Journal International*, *161*(2), 283–294, doi:10.1111/j.1365-246X.2005.02508.x.

Ni, S., V. F. Cormier, and D. V. Helmberger (2003), A Comparison of Synthetic Seismograms for 2D Structures: Semianalytical versus Numerical, *Bulletin of the Seismological Society of America*, *93*(6), 2752–2757, doi:10.1785/0120030011.

Panning, M., and B. Romanowicz (2006), A three-dimensional radially anisotropic model of shear velocity in the whole mantle, *Geophysical Journal International*, *167*(1), 361–379, doi:10.1111/j.1365-246X.2006.03100.x.

Ritsema, J., A. Deuss, H. J. Van Heijst, and J. H. Woodhouse (2011), S40RTS: a degree-40 shear-velocity model for the mantle from new Rayleigh wave dispersion, teleseismic traveltime and normal-mode splitting function measurements, *Geophysical Journal International*, *184*(3), 1223–1236, doi:10.1111/j.1365-246X.2010.04884.x.

Ritsema, J., S. Ni, D. V. Helmberger, and H. P. Crotwell (1998), Evidence for strong shear velocity reductions and velocity gradients in the lower mantle beneath Africa, *Geophysical Research Letters*, *25*(23), 4245–4248, doi:10.1029/1998GL900127.

Rost, S. (2002), Array seismology: Methods and applications, *Rev. Geophys.*, *40*(3), 1008, doi:10.1029/2000RG000100.

Rost, S., and P. S. Earle (2010), Identifying regions of strong scattering at the core–mantle boundary from analysis of PKKP precursor energy, *Earth and Planetary Science Letters*, *297*(3-4), 616–626, doi:10.1016/j.epsl.2010.07.014.

Rost, S., E. J. Garnero, and Q. Williams (2008), Seismic array detection of subducted oceanic crust in the lower mantle, *Journal of Geophysical Research: Solid Earth (1978–2012)*, *113*(B6), doi:10.1029/2007JB005263.

Rychert, C. A., and P. M. Shearer (2010), Resolving crustal thickness using SS waveform stacks, *Geophysical Journal International*, *180*(3), 1128–1137, doi:10.1111/j.1365-246X.2009.04497.x.

Song, T.-R. A., D. V. Helmberger, and S. P. Grand (2004), Low-velocity zone atop the 410-km seismic discontinuity in the northwestern United States, *Nature*, *427*(6974), 530–533, doi:10.1038/nature02231.

Tao, K., S. P. Grand, and F. Niu (2017), Full-waveform inversion of triplicated data using a normalized-correlation-coefficient-based misfit function, *Geophysical Journal International*, *210*(3), 1517–1524, doi:10.1093/gji/ggx249.

Chapter 5.

Artemieva, I. M., and W. D. Mooney (2001), Thermal thickness and evolution of Precambrian lithosphere: A global study, *Journal of Geophysical Research: Solid Earth (1978–2012)*, *106*(B8), 16387–16414, doi:10.1029/2000JB900439.

Auer, L., L. Boschi, T. W. Becker, T. N. Meyer, and D. Giardini (2014), Savani: A variable resolution whole-mantle model of anisotropic shear velocity variations based on multiple data sets, *Journal of Geophysical Research: Solid Earth*, *119*(4), 3006–3034, doi:10.1002/2013JB010773.

Becker, T., and L. Boschi (2002), A comparison of tomographic and geodynamic mantle models, *Geochem. Geophys. Geosyst.*, 1–48.

Crotwell, H. P., T. J. Owens, and J. Ritsema (1999), The TauP Toolkit: Flexible Seismic Travel-time and Ray-path Utilities, *Seismological Research Letters*, *70*(2), 154–160, doi:10.1785/gssrl.70.2.154.

Durand, S., E. Debayle, Y. Ricard, C. Zaroli, and S. Lambotte (2017), Confirmation of a change in the global shear velocity pattern at around 1000 km depth, *Geophysical Journal International*, *211*(3), 1628–1639, doi:10.1093/gji/ggx405.

Dziewonski, A. M., and D. L. Anderson (1981), Preliminary reference Earth model, *Physics of the Earth and Planetary Interiors*, *25*(4), 297–356, doi:10.1016/0031-9201(81)90046-7.

French, S. W., and B. Romanowicz (2015), Broad plumes rooted at the base of the Earth's mantle beneath major hotspots, *Nature*, *525*(7567), 95–99, doi:10.1038/nature14876.

Grand, S. P. (2002), Mantle shear–wave tomography and the fate of subducted slabs, *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *360*(1800), 2475–2491, doi:10.1098/rsta.2002.1077.

Garnero, E. J., and A. K. McNamara (2008), Structure and Dynamics of Earth's Lower Mantle, *Science*, *320*(5876), 626–628, doi:10.1126/science.1148028.

Garnero, E. J., A. K. McNamara, and S.-H. Shim (2016), Continent-sized anomalous zones with low seismic velocity at the base of Earth's mantle, *Nature Geosci*, *9*(7), 481–489, doi:10.1038/ngeo2733.

He, Y., L. Wen, and T. Zheng (2006), Geographic boundary and shear wave velocity structure of the “Pacific anomaly” near the core–mantle boundary beneath western Pacific, *Earth and Planetary Science Letters*, *244*(1-2), 302–314, doi:10.1016/j.epsl.2006.02.007.

He, Y., and L. Wen (2012), Geographic boundary of the “Pacific Anomaly” and its geometry and transitional structure in the north, *Journal of Geophysical Research: Solid Earth (1978–2012)*, *117*(B9), doi:10.1029/2012JB009436.

Houser, C., G. Masters, P. Shearer, and G. Laske (2008), Shear and compressional velocity models of the mantle from cluster analysis of long-period waveforms, *Geophysical Journal International*, *174*(1), 195–212, doi:10.1111/j.1365-246X.2008.03763.x.

Kennett, B. L. N., and O. Gudmundsson (1996), Ellipticity corrections for seismic phases, *Geophysical Journal International*, *127*(1), 40–48, doi:10.1111/j.1365-246X.1996.tb01533.x.

Koelemeijer, P., J. Ritsema, A. Deuss, and H. J. Van Heijst (2016), SP12RTS: a degree-12 model of shear- and compressional-wave velocity for Earth's mantle, *Geophysical Journal International*, *204*(2), 1024–1039, doi:10.1093/gji/ggv481.

Kustowski, B., G. Ekström, and A. M. Dziewonski (2008), Anisotropic shear-wave velocity structure of the Earth's mantle: A global model, *Journal of Geophysical Research: Solid Earth (1978–2012)*, *113*(B6), B06306, doi:10.1029/2007JB005169.

Lai, H., E. J. Garnero, S. P. Grand, R. W. Porritt, and T. W. Becker (2019), Global Travel Time Data Set From Adaptive Empirical Wavelet Construction, *Geochem. Geophys. Geosyst.*, *20*(5), 2175–2198, doi:10.1029/2018GC007905.

Lekic, V., S. Cottaar, A. Dziewonski, and B. Romanowicz (2012), Cluster analysis of global lower mantle tomography: A new class of structure and implications for chemical heterogeneity, *Earth and Planetary Science Letters*, *357-358*, 68–77, doi:10.1016/j.epsl.2012.09.014.

Li, C., R. D. van der Hilst, E. R. Engdahl, and S. Burdick (2008), A new global model for P wave speed variations in Earth's mantle, *Geochem. Geophys. Geosyst.*, *9*(5), n/a–n/a, doi:10.1029/2007GC001806.

Li, M., A. K. McNamara, and E. J. Garnero (2014), Chemical complexity of hotspots caused by cycling oceanic crust through mantle reservoirs, *Nature Geosci*, *7*(5), 366–370, doi:10.1038/ngeo2120.

Mégnin, C., and B. Romanowicz (2000), The three-dimensional shear velocity structure of the mantle from the inversion of body, surface and higher-mode waveforms, *Geophysical Journal International*, *143*(3), 709–728, doi:10.1046/j.1365-246X.2000.00298.x.

Moulik, P., and G. Ekström (2014), An anisotropic shear velocity model of the Earth's mantle using normal modes, body waves, surface waves and long-period waveforms, *Geophysical Journal International*, *199*(3), 1713–1738, doi:10.1093/gji/ggu356.

Nelson, P. L., and S. P. Grand (2018), Lower-mantle plume beneath the Yellowstone hotspot revealed by core waves, *Nature Geoscience*, *11*(4), 280–284, doi:10.1038/s41561-018-0075-y.

Panning, M., and B. Romanowicz (2006), A three-dimensional radially anisotropic model of shear velocity in the whole mantle, *Geophysical Journal International*, *167*(1), 361–379, doi:10.1111/j.1365-246X.2006.03100.x.

Ritsema, J., A. Deuss, H. J. Van Heijst, and J. H. Woodhouse (2011), S40RTS: a degree-40 shear-velocity model for the mantle from new Rayleigh wave dispersion, teleseismic traveltime and normal-mode splitting function measurements, *Geophysical Journal International*, *184*(3), 1223–1236, doi:10.1111/j.1365-246X.2010.04884.x.

Romanowicz, B. (1991), Seismic Tomography of The Earth`s Mantle, *Annu. Rev. Earth Planet. Sci.*, 1–23.

Simmons, N. A., A. M. Forte, L. Boschi, and S. P. Grand (2010), GyPSuM: A joint tomographic model of mantle density and seismic wave speeds, *J. Geophys. Res.*, *115*(B12), B12310, doi:10.1029/2010JB007631.

Sun, D., D. Helmberger, S. Ni, and D. Bower (2009), Direct measures of lateral velocity variation in the deep Earth, *J. Geophys. Res.*, *114*(B5), B05303, doi:10.1029/2008JB005873.

Suzuki, Y., K. Kawai, R. J. Geller, A. F. E. Borgeaud, and K. Konishi (2016), Waveform inversion for 3-D S-velocity structure of D″ beneath the Northern Pacific: possible evidence for a remnant slab and a passive plume, *Earth, Planets and Space 2016 68:1*, *68*(1), 198, doi:10.1186/s40623-016-0576-0.

Wang, Y. (2004), Mapping the geometry and geographic distribution of a very low velocity province at the base of the Earth's mantle, *J. Geophys. Res.*, *109*(B10), B10305–18, doi:10.1029/2003JB002674.